National Library of Canada
Bibliothèque nationale du Canada

Canadian Theses Service
Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, tests publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

NL-339 (r. 88/04)

Canada

THE UNIVERSITY OF ALBERTA

AN EVALUATION OF QUIZMASTER: A COMPUTER ASSISTED TESTING PROGRAM FOR INDUSTRIAL ARTS.

BY

© WALTER SCOTT LATTA

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF EDUCATION

IN

VOCATIONAL EDUCATION

DEPARTMENT OF INDUSTRIAL AND VOCATIONAL EDUCATION

EDMONTON, ALBERTA

FALL 1988

Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

The author (copyright owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-45519-5

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: WALTER SCOTT LATTA

TITLE OF THESIS: AN EVALUATION OF QUIZMASTER: A COMPUTER ASSISTED TESTING PROGRAM FOR INDUSTRIAL ARTS

DEGREE: MASTER OF EDUCATION

YEAR THIS DEGREE GRANTED: 1988

Permission is hereby granted to THE UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

W. S. Latta
(Student's signature)

279 Hillcrest Place
(Student's permanent address)
Edmonton, Alberta
T5R 5X6

Date: July 5, 1988

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled: AN EVALUATION OF QUIZMASTER: A COMPUTER ASSISTED TESTING PROGRAM FOR INDUSTRIAL ARTS

submitted by WALTER SCOTT LATTA

in partial fulfillment of the requirements for the degree of MASTER OF EDUCATION

in VOCATIONAL EDUCATION.

J.F. Dyck
.....
Supervisor

A.K. Deane
.....

Douglas V. Parker
.....

.....

.....

Date: June 27, 1988

Abstract

The primary purpose of this study was to determine if the microcomputer assisted testing program, entitled QUIZMASTER, was acceptable to Industrial Arts teachers and their students. A field test of the program was carried out in four junior high schools and two senior high schools. Since QUIZMASTER was still under development the field test was a formative evaluation of the program. Five teachers and 244 students participated.

The field test uncovered a number of problems with the computer program. These problems fell into three basic categories: 1) errors with the program code, 2) a lack of clarity in some of the program's messages and menu statements, and 3) program functions which needed greater depth and/or capacity.

It appears, from the data collected, that QUIZMASTER takes some time to learn to use. Once this hurdle is overcome the results show that the program was acceptable to those participating in the field test. A large majority of students indicated that they liked using a computer to take a test and preferred it over the traditional paper and pencil test methods. All of the teachers participating in the field test found that the program saved them time in the classroom. They believed that the using the program helped the students to learn the course content. They

indicated that they would use the program in their classrooms.

The field test did reveal some problems with the computer program. More importantly it elicited a number of suggestions on how to improve the program. The results of the study were positive enough to encourage further development of QUIZMASTER.

Acknowledgements

This thesis was possible only because a number of people cared enough to spend a significant amount of their time to try out QUIZMASTER and provide the feedback that was necessary to improve the program. In particular I would like to thank Messrs. Don Blackwell, Ray Guenette, Merne Hrycun, Terry Kent, Gary Kroy, Gordon Skeels, and Dan Wiles. These people took the time to use QUIZMASTER in their classrooms and were good enough to allow me into their classes to talk with their students. Without their considerable time and effort it would not have been possible to complete the evaluation of QUIZMASTER.

Thanks go the my thesis evaluation committee, Dr. J. F. D. Ilott, Dr. D. Parker, and Prof. A. K. Deane who took the time to review my work and suggest ways to improve it.

Special thanks go to my supervisor, Dr. J. F. D. Ilott, who allowed me to work in my own way but was always there to provide help and advice when it was needed most.

TABLE OF CONTENTS

CHAPTER		PAGE
1.	THE PROBLEM	1
	Introduction	1
	Background to the Problem	1
	The Computer Program	1
	Description of the Computer Program	2
	Evaluating the Program	5
	Statement of the Problem	6
	Definition of Terms	8
	General Terms	8
	Quizmaster Terms	8
	Research Questions	9
	Computer Program Code	9
	Computer Program Format and Content	9
	Computer Program Acceptability	10
2.	REVIEW OF THE LITERATURE	11
	Introduction	11
	Test Theory	11
	Achievement Testing	11
	Reasons for Testing	13
	Interpreting Test Results	15
	Scoring Objectivity	16
	Using Microcomputers in Testing	17
	Interactive Testing	19
	Establishing Computer Program Evaluation Criteria	22

Evaluation Procedures 25
Summary 26
3. METHODOLOGY 28
Introduction 28
The Model 29
Formative Evaluation 30
Procedure 32
Developing the Prototype System 32
Pilot Testing 32
Field Testing 35
Selecting Field Test Sites 36
Collection of Data 38
Collecting Data from the Students 38
Collecting Data from the Teachers 39
Informal Sources of Data 39
4. ANALYSIS AND INTERPRETATION OF DATA 41
Introduction 41
Referencing the Data 42
Examining the Data 42
Answering the Research Questions 43
Computer Program Code 43
Anticipated Results 44
Likert Scale Results 44
Identifying the Errors 45
Computer Program Format and Content 47
Anticipated Results: Ease of Use 47

Likert Scale Results: Ease of Use 48
Identifying the Problems: Ease of Use 50
Anticipated Results: Test Format 52
Likert Scale Results: Test Format 53
Identifying Test Format Problems 55
Anticipated Results: Program Completeness . . . 57
Likert Scale Results: Program Completeness . . 58
Identifying the Shortcomings 59
Computer Program Acceptability 60
Anticipated Results 60
Likert Scale Results 61
Discussion of Results 64
5. CONCLUSIONS AND RECOMMENDATIONS 68
Purpose of the Study 68
Summary of the Results 69
Computer Program Code 69
Computer Program Format and Content 70
Computer Program Acceptability 72
Modifying the Program 73
Fixing the Code Errors 74
Conclusions 75
Recommendations 77
BIBLIOGRAPHY 79

APPENDIX 1. FIELD TEST QUESTIONNAIRES AND INTERVIEW SCHEDULES . . 83

APPENDIX 2. PILOT TEST QUESTIONNAIRE AND PROGRAM FAULT REPORT . . 91

APPENDIX 3 SUMMARY OF DATA 95

APPENDIX 4. QUIZMASTER PROGRAM DESCRIPTION 122

LIST OF TABLES

Table 4-1. Computer Program Code: Questions and Data Summary . . 45

Table 4-2. Likert Data Summary: Ease of Use 48

Table 4-3. Likert Data Summary: Test Format 54

Table 4-4. Likert Data Summary: Program Completeness 59

Table 4-5. Likert Data Summary: Program Acceptability 62

LIST OF FIGURES

Figure 3-1. Comparison of Formative and Summative Evaluation . . 29

Figure 3-2. Major Activities in Formative Evaluation 30

Chapter 1: The Problem

Introduction

This thesis describes the evaluation of a computer assisted testing program, entitled QUIZMASTER, which was developed at the University of Alberta. The thesis is divided into five chapters. The first chapter describes the study, provides some background information, and outlines the research questions. Chapter 2 is a review of the literature relevant to achievement testing theory, the use of microcomputers in testing, and systems for evaluating microcomputer programs. The third chapter describes the methodology used in carrying out the study. Chapter 4 provides an analysis of the data that was collected during the study. A more complete summary of that data can be found in Appendix 3. The last chapter summarizes the data and makes some recommendations for further work in this area.

Background to the Problem

The Computer Program

Before proceeding with any detailed discussion of the evaluation process it is useful to examine, briefly, the computer program which is the object of this study. The computer program, entitled QUIZMASTER, is a microcomputer assisted testing program developed within the Department of Industrial and Vocational Education at the University of Alberta by J. F. D. Ildtt and W. S. Latta.

QUIZMASTER was intended to serve as a laboratory management aid for Industrial Education teachers. The multiple activity nature of the Industrial Arts program implies a classroom management structure where students complete assigned learning tasks on a rotational basis. As a consequence the teacher must repeat lessons, demonstrations and evaluations many times during the school year.

Student evaluation is an area of major concern for the Industrial Arts teacher. It is necessary to have the students complete a number of safety and theory quizzes during the course of each rotation. Since these tests are normally written on an individual basis the act of test administration directs the teacher's attention away from the task of laboratory supervision and instruction. The use of traditional paper and pencil tests often leads to a problem with test security. QUIZMASTER is intended to be a program which can be operated by the students with a minimum of teacher assistance. The purpose is to free the teacher from the routine tasks of invigilating and marking tests. QUIZMASTER is capable of selecting questions at random from the test bank, thus relieving the problem of test security.

Description of the Computer Program

QUIZMASTER's organization is centred around the Unit. The term Unit as it is used here is analogous to test or

unit test. QUIZMASTER can handle up to thirty Units. The Unit structure is fundamental to the way that QUIZMASTER presents a test. The Unit is made up of a number of components that can be assembled in a variety of ways to make the test that is presented to the student.

Each Unit is broken down into Topics. Each Topic represents one specific topic or objective to be tested. Consider, for a moment, the processes involved in making up a test for Driver Education. The test name, or in the case of QUIZMASTER, the Unit name, might be "Rules of the Road". The test could contain one Question each from such Topics as signalling for turns, speed limits, changing lanes, and so on. There is provision for an item pool of four Questions in each Topic. In order to produce a different test each time a Unit is run, QUIZMASTER selects one Question at random from each Topic. Since a Unit can accommodate up to twenty Topics, each test can have a maximum of twenty Questions. A Unit can contain a maximum of eighty Questions, four in each of the twenty Topics.

QUIZMASTER has full editing features which allow the teacher to enter his or her own tests or to modify the tests provided with the program. The latter option makes it possible to amend tests to make them suitable for the teacher's course of studies. QUIZMASTER is able to handle multiple choice, true-false, and completion type Questions.

The test presentation includes the use of two optional introductory screens, the Unit Objective and the Student Introduction. These two screens can be used for a variety of purposes. The intent of the former is to inform the student of the learning objectives established for the test. The latter of the two screens is intended to inform the student of any special requirements that must be met before the Unit can be completed.

The teacher also has the option of entering resource lists which can provide the student with sources of information for review purposes.

An important aspect of computer assisted testing is the ability to provide the student with appropriate information on success or failure for each test item. The teacher has the option of entering both Mastery and Non-Mastery Feedback for each Question as well as Mastery and Non-Mastery Feedback for the Unit.

A number of other features are under teacher control. These include setting the mastery level or pass mark, determining whether the Feedback statements should be displayed during the test or be delayed until the test review, and controlling student access to any Unit in order to control course sequencing.

QUIZMASTER includes a number of record keeping functions. The student's performance on a Unit, together with a record of which questions were presented, is

maintained for the first and last attempts at a Unit for each student. Records are also kept on the responses for each Question so that an item difficulty index can be computed and inappropriate distractors in multiple choice items can be identified.

Paper and pencil tests can be generated by the computer either on a random basis or under teacher control.

A more complete description of QUIZMASTER can be found in Appendix 4.

Evaluating the Program

When developing educational computer software it is important that the program development be accompanied by a logical system of evaluation. Kearsley (1983, p. 107) presents a workable format for computer program development and formative evaluation. He describes the developmental process as occurring in three phases. During the first, or design phase the evaluation takes the form of a review of the program specifications by experts in subject matter and instructional design. The second, or developmental, phase is evaluated through pilot testing with small groups. This provides the opportunity to try out ideas and evaluate the effectiveness of the software. The third phase of development and evaluation is the implementation phase. This tests the effectiveness of the software in an operational setting. It is the third or implementation phase which is the primary focus of this study. Kearsley

advocates a summative evaluation be carried out upon completion of the foregoing three phases of formative evaluation. The summative evaluation is beyond the scope of this study.

During the early phases of program development QUIZMASTER underwent phase one and phase two, the review and pilot testing phases, of the evaluation procedure just described. The first phase consisted primarily of discussing the program specifications with educators in the Department of Industrial and Vocational Education as well as with practicing teachers. Once the program code had been written, two different pilot tests were conducted. Both used undergraduate students enrolled in teaching methods courses. The first pilot test was limited in that the students only used the program to generate Unit tests. The second pilot study had the university students generate Unit tests then use the computer to assist them in evaluating junior high school students in a practice teaching situation.

Statement of the Problem

The primary focus of this study was to evaluate the acceptability of the microcomputer assisted testing program, QUIZMASTER, to industrial education teachers and students. In order to accomplish this it was necessary to seek answers to three broad questions: 1) Is QUIZMASTER's computer code correct and complete? 2) Is QUIZMASTER'S

format and content acceptable? 3) Is QUIZMASTER acceptable to teachers and students?

Is QUIZMASTER's computer code correct and complete? For a computer program to be acceptable, in any sense, it is critical that all syntax and logic errors be identified and corrected. Although extensive testing had been carried out on QUIZMASTER the possibility existed, considering the size and complexity of the program, that some code errors remained.

Is QUIZMASTER'S format and content acceptable? The point of this question was to determine if QUIZMASTER's format and features were consistent with what teachers would expect in a computer assisted testing program. Up to the point of the field test, the program specifications had been determined through discussion with people familiar with computer assisted testing and through examining other computer testing programs to identify the more desirable features. There was no guarantee that Industrial Arts teachers would find the computer program format acceptable or that all the desirable features had been incorporated.

Is QUIZMASTER acceptable to teachers and students? If QUIZMASTER is to be useful in the classroom both the students and the teachers should see some value in the program and feel comfortable with using it.

Definition of Terms

General Terms

CAT. Computer Assisted Testing. See Computer based Testing.

CBT. Computer Based Testing. Also referred to as Computer Assisted Testing (CAT). CBT "is the use of the computer to generate, administer, score, or analyze tests" (Kearsley, 1983, p. 20).

Multiple Activity Laboratory. "A laboratory where three or more activities are in progress at the same time." (Alberta Education, undated)

QUIZMASTER Terms

QUIZMASTER. A CAT program designed for use on the Apple II series of computers. In addition to administering tests it maintains student records, allows control of student progress, is capable of providing appropriate Feedback statements, and enables the teacher to author quizzes.

Unit. In QUIZMASTER the term Unit is used to refer to a complete test made up of many components including Topics, Questions, Resource lists, Mastery Feedback statements, and Non-Mastery Feedback statements.

Topic. In QUIZMASTER a Topic represents an item pool of four questions. A Unit may contain twenty Topics. When a student uses QUIZMASTER to take a test the test is made up by selecting one question from each of the existing topics.

Question. In QUIZMASTER the term Question refers to one of the questions in a topic. The Question may be multiple choice, true-false or completion type.

Mastery Feedback. In QUIZMASTER Mastery Feedback statements may be presented on those occasions when a student makes a correct response to a Question. The intent is to confirm to the student that the question was answered correctly and then provide additional information as necessary. The content of the statement is determined by the test author.

Non-Mastery Feedback. In QUIZMASTER, Non-Mastery Feedback statements may be presented on those occasions when a student makes an incorrect response to a Question. The content of the statement is determined by the test author.

Research Questions

The research questions posed by this study are listed below.

Computer Program code

1. Does the computer program cease operation at unexpected times?
2. Does the computer program respond to commands and data input in the expected manner?

Computer Program Format and Content

1. Is the computer program user friendly?
2. Are the menu functions clear?

3. Is the screen format acceptable?
4. Have any important functions been omitted?
5. Are there any superfluous functions?
6. Is the test format acceptable?
7. Are premade tests desirable?

Computer Program Acceptability

1. Will students accept this form of testing?
2. Do teachers see the computer program as a time saving device?
3. Is there sufficient value in the computer program for teachers to create their own test banks?

Chapter 2: Review of the Literature

Introduction

This review of literature has several purposes. First there is a need to examine established achievement testing theory. This is necessary to the evaluation process since QUIZMASTER, as a computer assisted testing program, should conform, in its major aspects, to established theory. The QUIZMASTER program is described in Appendix 4. The second part of this review examines how microcomputers can be used in the testing process. The criteria used to evaluate computer programs and literature concerning the evaluation of computer software are also reviewed.

Test Theory

The following discussion focuses on achievement testing. It attempts to look at some of the reasons for administering tests and some of the methods of interpreting the results of achievement tests. The point of the discussion is to illustrate one type of testing that can be successfully carried on with the use of an interactive microcomputer testing program such as QUIZMASTER.

Achievement Testing

Achievement testing plays a very important role in the educational process. Gronlund (1982) sees achievement testing as "the most widely used method of assessing pupil achievement in classroom instruction, and it is an indispensable procedure in individualized and programmed

instruction" (p. 1). He defines an achievement test as "a systematic procedure for determining the amount a student has learned" (p. 1). Popham (1981) defines achievement testing in relation to cognitive and psychomotor behaviors.

> Tests in the cognitive or psychomotor realm are often focused on an examinee's attainments at a given point in time; these tests are usually referred to as achievement tests. (p. 33)

The "point in time" referred to by Popham (1981) is important because it, in part, determines the reasons for testing and the form that a test may take. The effects of the timing of a test are discussed later in this chapter in the section "Reasons for Testing".

Gronlund (1982) puts forward and discusses six basic principles of achievement testing. Those principles, which are listed below, provide a good description of achievement testing and how it can be used.

1. Achievement tests should measure clearly defined learning outcomes that are in harmony with the instructional objectives.
2. Achievement tests should measure a representative sample of the learning tasks included in the instruction.
3. Achievement tests should include the types of test items that are most appropriate in classroom testing.

4. Achievement tests should fit the particular uses that will be made of the results.
5. Achievement tests should be as reliable as possible and should then be interpreted with caution.
6. Achievement tests should improve student learning. (pp. 8 - 13)

Reasons for Testing

It is useful to look at the testing process in terms of the evaluation decisions that the teacher needs to make. Gronlund (1982) suggests three types of decisions that can be aided by testing. First there are decisions made at the beginning of instruction. These are decisions related to readiness and placement. The second type of decisions that are made occur during the course of instruction. These decisions are related to student progress and are often used to identify learning problems. The last type of decision is made at the end of instruction and is often used to form the basis for student promotion (p. 2). Bloom, Hastings, & Madaus (1971) devote several chapters to discussing the characteristics of these testing decisions and of the tests best suited to meet the needs. These testing decisions translate into three basic classes of testing; 1) formative evaluation, 2) summative evaluation, and 3) evaluation for diagnosis.

Formative evaluation. Formative evaluation is normally carried out during the course of instruction for the

purpose of monitoring student progress and providing feedback regarding the student's strengths and weaknesses. Its purpose is to improve the student's learning. As Bloom et al. (1971) point out formative evaluation "points to areas of needed remediation so that immediately subsequent instruction and study can be made more pertinent and beneficial" (p. 20). The tests used are usually limited in scope to a single unit of instruction (Gronlund, 1982, p. 3).

Summative evaluation. Summative evaluation is normally carried out at the end of instruction or at intervals during the course of instruction. It is done with the idea of providing students with grades or certification. The tests used usually cover a wide range of objectives and tend to sample the learning tasks rather than cover them all in detail (Bloom et al., 1971, p. 20; Gronlund, 1982, p. 4).

Evaluation for diagnosis. Diagnostic tests cover a wide range of topics and often report a number of subscale marks in addition to an overall mark. Tests given prior to the start of instruction normally have placement as a primary goal. Diagnostic tests can be used during instruction in an attempt to identify causes of difficulties that a student may have in learning the material (Bloom et al., 1971, p. 87).

The form that a test may take will depend upon when the test is administered and the relative importance of the test. Testing carried out for formative and diagnostic purposes would likely have a higher degree of feedback for the student. The timing of any feedback to the students could be affected by the perceived importance of a test. As Alessi and Trollip (1985) point out; "The general Practice is to delay feedback on more important tests until the entire test has been completed. On the other hand, less formal classroom tests often provide feedback immediately" (p. 246).

Interpreting Test Results

According to Gronlund (1982, p.108) tests can be interpreted in two basic ways; norm-referenced interpretation or criterion-referenced interpretation. In the first case the student's performance is measured against other members of the class. In the second case performance is measured against the objectives for the course of studies. Popham (1981) provides working definitions for both types of test interpretation.

> A norm referenced test is used to ascertain an individual's status with respect to the performance of other individuals on that test. (p. 26)
>
> A criterion referenced test is used to ascertain an individual's status with respect to a defined behavioral domain. (p. 27)

The difference between norm and criterion referencing is in the interpretation of the test results and not in the test itself. The different interpretations, however, do lead to differences in test construction. Gronlund (1982, pp. 14-16) provides a useful summary comparing the two approaches and highlights how norm referenced tests and criterion referenced tests would differ in content. He gives a more detailed description on test construction in his book Measurement and evaluation in teaching (1981, chap. 5)

In order to provide the proper emphasis, item selection becomes quite important. Although the selection procedures vary, both types of testing require item selection according to a predetermined plan. It would seem that a CAT program that selects items in a structured manner as opposed to pure random selection would be preferred.

Scoring Objectivity

One of the benefits of using a microcomputer to administer tests is the consistent, and presumably objective manner in which it scores the tests. The consistency with which a computer scores tests, however, does not guarantee objectivity. Certainly, as Alessi & Trollip (1985, p. 98) point out, essay questions cannot even be properly evaluated by computers at this time. Completion items that can be answered with a single word or

phrase can have several different correct responses and different teachers may accept a different range of answers for the same question. (Bloom et al., 1971, p. 83) Even the so called objective multiple choice items can cause problems in scoring. This is particularly true when students are instructed to choose the best response to a question. It is possible, with this type of question, for competent teachers to have honest disagreement about which response should be keyed as the correct answer. (Bloom et al., 1971, p. 82) To overcome this type of problem it is important that the test developer construct the test items in such a way that there is only one correct response to objective items and a clear, limited range of answers to completion items. Gronlund (1982) devotes several chapters to the problem of preparing various types of test questions.

Using Microcomputers in Testing

If microcomputers are to be used in assessing achievement and instruction it must be established that the microcomputer will contribute in some way to the evaluation process. It is useful to look at the testing process, break it down into a number of steps, then attempt to determine what contributions, if any that the microcomputer can make to the process. Both Nelson (1984, p. 22) and Roeber (1984, p. 27) look at the testing process and attempt to analyze it in this way.

Nelson (1984, p 22) offers the following list of steps to describe the testing process: 1) creating tests, 2) giving tests, 3) scoring tests, 4) determining test quality, 5) reporting results, 6) recording results in class grade book, 7) reflecting on the quality of instruction, 8) deriving final grade at term's end, and 9) reporting final grades to students and central administration. Nelson (1984) contends that steps 3, 5, 6, 8, and 9 "can be considerably assisted by using a computer" (p. 22). Roeber (1984) sees a similar use for computers.

Grade book type programs that support the processes indicated in steps 5, 6, 8, and 9 are easy to find in many software catalogs. Hsu & Nitko (1983, p. 24) offer a number of examples. In addition templates are available that will permit the use of most spreadsheet type programs to be used for grade book functions. Flast & Flast (1986, p. 59) offer a simple example of a spreadsheet application in their book Appleworks Applications.

The process of scoring paper and pencil tests, if it is to be efficient, requires special optical scoring equipment to input data. (Roeber, 1984, p. 27) When microcomputers are used to administer tests, however, the microcomputer can score the test and record the student's mark. It is possible, with microcomputer administered tests, to store other types of information. For example, if time is important to the evaluation process the student's

rate of progress can be monitored, and this information can be integrated into the student's mark. (Ward, 1984, p.16)

The process for actually creating tests is a problem area for some teachers. Hambleton (1984, p. 11) cites a shortage of time to develop tests as being a common cause for complaint. Poor facilities for producing tests and the teachers's own limited skills are also given as problem areas for many teachers.

The use of computer storage for good quality questions in a properly coded item bank can provide some relief for teachers. When tied to some type of word processor or test printing package a good item bank can make the test assembly process easier. (Hambleton, Anderson, and Murray, 1983, p. 68; Roeber, 1984, p. 27)

Interactive Testing

Using the computer to create and administer tests has several benefits. The chance of errors in scoring the test are reduced. Students can also receive their results sooner. (Hambleton et al., 1983, p. 76) More frequent testing is possible using the computer. Some teachers have found this to be a distinct advantage. (Noble, 1981, p. 386) Student records are maintained by the computer allowing the teacher almost instant access to student marks as well as more detailed profiles on student progress.

Many separate programs exist for conducting an item analysis on multiple choice tests. Two such programs are

discussed by Hsu & Nitko. (1983, p. 18) These programs allow the teacher to validate the test item bank and to assign difficulty levels to each question. Nelson (1984, p. 24) and Hambleton (1984, p. 11) both discuss the value of the ability to assess test item quality. The item analysis feature can be easily built into a computer testing program, thus allowing an ongoing analysis.

Another area of concern in any type of testing is test security. The computer's ability to select items from an item bank in a variety of ways reduces the security problem. (Hambleton et al., 1983, p.70) The ability to create parallel forms of a test provides for the repeated testing that is characteristic of mastery learning and of individualized instruction. (Millman & Arter, 1984, p. 317) The ability to select items in a random manner appears to bring another benefit.

> As one might expect, students will discuss questions with each other prior to an examination or quiz. However, since there are a number of questions for each objective and the options are scrambled, students tend to discuss underlying concepts rather than answers to specific questions. (Dunkleberger & Heikkinen, 1983, p. 35)

The use of microcomputers to aid in the construction of tests has a number of advantages. The use of a validated item bank and a word processor makes it possible to

generate a number of parallel forms of test more quickly and with fewer errors. A variety of computer programs are available to assist in recording and reporting student marks. Item analysis programs make it possible to gather and organize the data necessary to improve the quality and validate the items in an item bank. Interactive testing programs make it possible to have a test retest capability without compromising test security and/or test quality. It is possible to integrate any or all of the aforementioned functions into a single interactive program.

Ward (1984) discusses some of the advantages that can be realized by using the microcomputer to individualize student assessment. He points out that; "Efficient use of the equipment calls for a more flexible scheduling; students can be allowed to take tests at virtually any time" (p. 16). It also removes the time constraints imposed by group testing. There is no longer any need to require all students to complete the test within the same time period. If time is an important consideration, it can be monitored and recorded by the computer. With computers it is possible to employ different assessment strategies by providing immediate feedback on the correctness of each response or to allow a student to continue to respond until the correct answer is obtained. Millman (1984, p. 20) in critiquing Ward's work, supports the views that Ward advances on the individualization of testing.

Establishing Computer Program Evaluation Criteria

There are two sides to establishing computer program evaluation criteria. First are the operational and administrative considerations. These cover a wide variety of concerns including such concepts as machine compatibility, presentation format, use of graphics, program features, and the vendor's sales and warranty procedures. The second and possibly more important side deals directly with learning issues.

One way of dealing with both the operational and administrative criteria, as well as with the learning issues is to prepare an evaluation checklist or alternatively a list of questions to be considered when evaluating a computer program. There are many examples of these kinds of lists in the literature. The Clearinghouse of Alberta Education (1984) has published an extensive list of questions to be used by their computer program evaluators. Lathrop & Goodson (1983) devote four chapters of their book to the evaluation process and provide several lists of questions dealing with various aspects of the evaluation process. Hsu & Nitko (1983, p. 27) offer a comprehensive system which they used to evaluate a number of microcomputer testing programs. Kenneth Krause (1984, p.24) puts forward a list of questions that he has found useful in evaluating reading software. Other lists of

evaluation criteria are advanced by Jelden (1982, p. 32) and Senn (1983, p. 317).

Caution must be exercised when using any checklist. As Miller and Burnett (1986, pp. 159-160) point out when criticizing Krause's set of guidelines, a number of problems can arise. One concern is that an imbalance in the checklist may occur which will overemphasize the operational and administrative concerns and down play the theoretical concerns related to the learning issues. Rawitsch (1983, p. 332) also holds this point of view. Another problem area lies in the wording of items in the checklist. Poor wording may inadvertently restrict the number of software packages that receive a positive evaluation. This could lead to the rejection of otherwise valuable software packages. A third area of concern is that software designers may tailor their programs to meet established evaluation criterion in order to ensure a positive evaluation. Checklists and lists of questions are useful devices to use when evaluating educational software. Their use ensures that the most important concerns are dealt with in the evaluation process. It is important, however, to keep in mind the shortcomings of lists in order to prevent overemphasizing minor points and to avoid rejecting software packages for the wrong reasons.

The evaluation of QUIZMASTER differs from most of the evaluations described in the referenced articles in that it

involves the evaluation of a CAT program under development and not the evaluation of a finished product. The evaluation of a program under development is essentially formative evaluation. Most of the articles reviewed dealt with summative evaluation procedures. Consequently none of the checklists or lists of questions reviewed fit the needs of the formative evaluation of QUIZMASTER. The exercise of reviewing this literature, however, proved useful. Without going so far as tailoring software to meet current evaluation standards, it is possible to determine the more desirable qualities for CAT software. Those qualities seem to be:

1. The program should be able to access a question item bank choosing items from that bank in some random fashion. (Dunkleberger & Heikkinen, 1983, p. 35; Hambleton et al., 1983, p. 76)
2. The program should be capable of storing data on each question so that an item analysis can be carried out for the purpose of validating the test items. (Hambleton et al., 1983, p. 76; Jelden, 1982, p. 32; Senn, 1983, p. 317)
3. Some method must exist within the computer program for making changes to items in the question bank. (Hambleton et al., 1983, p. 76; Hambleton, 1984, p. 12; Hsu & Sadock, 1985, p. 6; Jelden, 1982, p. 32)

4. The program should provide some form of feedback to the student regarding his progress. (Cartwright & Derevensky, 1976, p. 318; Hambleton et al., 1983, p. 76; Mizokawa & Hamlin, 1984, p. 16)
5. The program should be capable of maintaining progress records on each student and should be able to display both student and class progress profiles. (Hambleton et al., 1983, p. 76; Mizokawa & Hamlin, 1984, p.16; Senn, 1983, p. 317)
6. The teacher should be able to exert some form of control over the program in regard to test difficulty and student progression. (Hambleton et al. 1983, p. 70; Noble, 1981, p. 385)
7. The program should be capable of generating a number of different types of tests including pretests and post tests. (Jelden, 1982, p. 32)

Evaluation Procedures

The evaluation process used to assess a computer program under development is a formative evaluation process. It is done with the idea of improving the quality of the computer program. Kearsley (1983) differentiates between formative and summative evaluation this way:

> In contrast to the purpose of formative evaluation, which is to improve the efficiency or effectiveness of a system or training program, the purpose of summative evaluation is to determine the efficiency or

effectiveness and to make decisions based on this assessment. (p. 152)

Formative evaluation models have one thing in common, they are cyclical. They provide a mechanism for collecting information, analyzing that information, making changes based on the information collected, then restarting the information collection process. Harris, Bell, and Carter (1981, p. 1.1.2/ 2) describe one such process. This model was used by Layman & Hall (1986) to evaluate a computer based teaching package. Another model was developed by Kearsley (1983, p. 146) specifically for the purpose of evaluating computer based training systems. Kearsley's model, which is described more completely in Chapter 3 was used as the model for the evaluation of QUIZMASTER.

Summary

When examining any CAT program it is useful to be able to describe it in terms of some testing theory in order to describe the program's capabilities. When QUIZMASTER was conceived its purpose was to relieve some of the problems associated with achievement testing in the multiple activity Industrial Arts program. Since the testing needs were primarily for formative evaluations, QUIZMASTER is strongest for this type of testing. Its capabilities can, with some limitations, be extended into summative and diagnostic evaluation activities.

There is little point in using microcomputers in the testing process if they do not contribute something in the way of time savings or improvements in the testing situation. Microcomputers can save time in the preparation of tests. The number of errors in the test can be reduced and the quality of the items monitored and improved. Interactive testing programs help reduce the chance of errors in scoring tests and make it possible to increase the frequency of tests. Test security is generally improved by the random selection procedures most CAT programs use to assemble tests. There is also some indication that the random selection feature may improve the learning situation. The introduction of the microcomputer into the classroom makes it convenient to individualize assessment and to employ different assessment strategies.

There are many methods available for use when evaluating CAT programs. It is relatively easy to find checklists and lists of criteria that are deemed desirable. Perusing these lists makes it possible to come up with a list of common features that most authors deem to be important.

To make the evaluation process complete, all of the foregoing ideas can be brought together inside some evaluation model so that judgments about a CAT program can be made.

Chapter 3: Methodology

Introduction

The search for a method to evaluate educational computer programs leads one into the domain of program and course evaluation methods. These methods can be modified so that evaluation of computer programs is possible. One such method is suggested by Kearsley (1983) in his book Computer-Based Training: A Guide to Selection and Implementation. In his system he groups the various evaluation activities into two types of evaluation: formative and summative. Anderson (1985, p. 200) takes a similar approach to the problem.

It is important to draw a distinction between the purposes of formative and summative evaluation. With formative evaluation the purpose is to gather information to identify ways in which a program can be improved. To this end the evaluation activities would normally be carried out during the developmental stages. Summative evaluation activities are conducted for the purpose of gathering information to aid in making final judgments or decisions. (Anderson, 1985; Kearsley, 1983) According to Kearsley (1983), in discussing the summative evaluation of computer based training (CBT) "the purpose of summative evaluation is to allow decision-makers to assess the value of a system in terms of alternative approaches (comparative), effects on the job performance (validation),

or cost/benefits" (p. 144). Figure 3-1 illustrates the essential differences between formative and summative evaluation.

	FORMATIVE	SUMMATIVE
PURPOSE	Improve Program	Adoption of Program
TYPE OF DATA	Program Performance	Competitive, Validation, Cost/Benefits
STATUS OF SYSTEM	Prototype	Operational
EVALUATOR	Usually Internal	Usually External
DECISION MAKER	Instructional Specialists	Management

Figure 3-1. Comparison of formative and summative evaluation. (Adapted from Kearsley, 1983, p. 145)

The Model

The system that Kearsley (1983) describes incorporates two different types of evaluation: formative and summative. Since the purpose of the study described in this report was to evaluate a prototype computer program, only the formative evaluation procedures are of interest. The formative evaluation structure put forth by Kearsley (1983) served as the basis for the evaluation described in this report.

Formative Evaluation

The formative evaluation process is illustrated in Figure 3-2. This figure shows the cyclical nature of the evaluation process. The model is essentially a linear progression which subjects a prototype system to a pilot test and a field test with the ultimate goal of producing an operational system. The formative evaluation component of the model makes it possible to evaluate each stage of the process as it occurs and then, if necessary, go back and make changes to the program being evaluated.

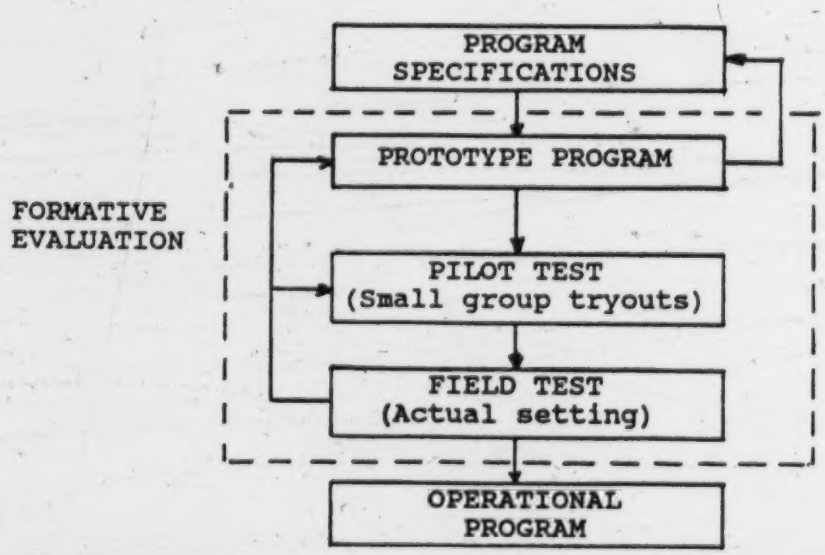

Figure 3-2. Major Activities in Formative Evaluation. (Adapted from Kearsley, 1983, p. 146)

The first test in the formative evaluation process is to subject a prototype system to a pilot test or small group tryouts. The purpose at this stage is to identify and

correct as many errors and problems as possible. During the pilot testing phase the process involved in making corrections is relatively simple because, as the model shows, the return path, to correct the system specifications, is short, and only a few people are involved. Even radical changes in the prototype system are easy to handle because the changes need only be explained to a few people. It seems unlikely that an ideal system would emerge from the pilot test phase but the system that does result will, in all probability, be a working system.

The field test phase is entered when the results of the pilot test are judged satisfactory. The field test is a more demanding test for two reasons. First, there is a greater potential for uncovering problems because there are more people involved. Secondly, the field test, because it takes place in an operational environment, is a better and more demanding test of the specifications and system design.

As was the case with the pilot test, in the field test there is provision to move back in the evaluation cycle and make the necessary changes to the system and/or system specifications. Because more people are likely to be involved and the system is being used for its intended purpose, the process involved in making changes and reevaluating the modified system is necessarily more difficult.

Procedure

Developing the Prototype system

For the purposes of this study the prototype system being evaluated was the microcomputer based testing program, QUIZMASTER, which is described in Appendix 4. The program was developed by J. F. D. Ilott and W. S. Latta at the University of Alberta. The initial motivation to develop the program was to provide a teaching tool which would be of assistance to Industrial Arts teachers in managing multiple activity laboratories. Advice was sought from several members of the Department of Industrial and Vocational Education during the developmental process. The early stages of testing, including pilot testing, was done by the test authors with the cooperation of other members of the Department of Industrial and Vocational Education. A brief description of the pilot test is given in the next section.

Pilot Testing

QUIZMASTER was pilot tested by a number of undergraduates registered in the Bachelor of Education program in Industrial Arts. These students were asked to use the program to create a QUIZMASTER Unit and if possible use the Unit in a practical situation. The pilot test was carried out in conjunction with the normal class activities associated with the Curriculum and Instruction course; EDIND 391. EDIND 391 is a course which provides an

opportunity for the undergraduate students to practice their teaching skills in a carefully supervised Industrial Arts multiple activity laboratory located on campus.

Aside from being convenient, using the EDIND 391 students and facilities for the pilot test had several advantages. First, it provided an opportunity for people with little or no computer experience to use the program. This was a good test of the program's ease of use and error handling capabilities. The EDIND 391 teaching facility has a number of Apple II computers thus solving the problem of access to the necessary hardware. A third advantage of using this class was that the possibility existed that the computer program could be used for its intended end purpose of evaluating students. Unfortunately only one or two individuals got far enough along with the project to actually use the program with the junior high school students.

The activities generated by using QUIZMASTER were within the objectives of the EDIND 391 course of studies. Although the students were participating in a pilot test of a computer program, they were doing so for course credit.

Prior to the start of the pilot test it was necessary to give introductory instruction in using microcomputers and in test item construction. The instruction on microcomputer operation was necessary since many of the students had no knowledge of computer operation.

Instruction in test item construction and testing procedures was consistent with the EDIND 391 course objectives.

Throughout this phase it was necessary to keep in close contact with the EDIND 391 students to assist them in computer operation and test item construction. It was assumed that there were still problems with QUIZMASTER. This turned out to be the case. Working closely with the students during the pilot test provided support for the EDIND 391 students in overcoming problems in computer operation and test item construction. To facilitate the identification of programming errors the computer program had a built in error message and the students were provided with a program fault report form on which they could detail any programming errors that they encountered. A sample of this form is shown in Appendix 2.

At the end of the pilot test the students were asked to complete a questionnaire similar to the one shown in Appendix 2. Information gleaned from this questionnaire, program fault reports, and ongoing discussions with the students were used to make a decisions on program changes in order to prepare QUIZMASTER for the field test. The questionnaire was also be evaluated at this time to determine if it was in fact yielding the desired information. Some changes to the questionnaire were necessary.

Field Testing

The field test was carried out in the Industrial Arts laboratories of six junior high schools and two senior high schools. There were only two criteria for teachers to participate in the field test. First it was necessary for them to have an Apple II+, Apple IIe, or Apple IIc computer located in the Industrial Arts laboratory. The second was that the teacher be willing to use the program to author at least one test and to use the program as a part of their student evaluation scheme.

Each teacher participating in the field test was given a familiarization session with QUIZMASTER prior to starting the field test activities. They were asked to generate at least one QUIZMASTER Unit and then use the program in their laboratories with at least one class through one class rotation. The actual timing of the use of QUIZMASTER in the classroom was determined by each teacher's laboratory management schedule. During the period of the field test, contact was maintained by telephone and visits were made to the schools whenever a teacher encountered problems that could not be handled over the telephone. It was necessary to make several trips to some of the participating schools.

At the end of the field test period each teacher was asked to complete a questionnaire similar to the teacher questionnaire illustrated in Appendix 1. Each teacher was interviewed to determine any reactions and problems not

covered in the questionnaire. The interview schedule is also shown in Appendix 1. At the participating teacher's discretion, a questionnaire was administered to those students who completed tests using QUIZMASTER and a number of students were interviewed. The student questionnaire and the student interview schedule are illustrated in Appendix 1.

Selecting Field Test Sites

The field test for QUIZMASTER was carried out in the Industrial Arts laboratories of the participating junior and senior high schools. The Industrial Arts program was selected for a number of reasons.

The Industrial Arts program, as it is implemented in the Province of Alberta, emphasizes individual progress through a multiple activity approach to teaching the course content. A multiple activity laboratory is defined in the Junior High School Curriculum Guide as "A laboratory where three or more activities are in progress at the same time" (Alberta Education, undated, p.7). This attribute was particularly important because QUIZMASTER was developed as an aid to individualizing instruction. One of the motivating factors in developing QUIZMASTER was to provide Industrial Arts teachers with some assistance in dealing with the complex problems of testing schedules and test security which result from the operation of the multiple activity program.

Another factor in choosing the Industrial Arts program was the availability of Apple II computers. A number of the laboratories were already equipped with these computers as a result of program needs associated with teaching the Graphic Communications, Electricity/Electronics, and Computer Modules.

One of the requests made of the participating teachers was that they use QUIZMASTER in the same way that it would be used if it was integrated into their normal evaluation process. The smaller size of the Industrial Arts classes made it possible to approach 100% student participation.

There is a natural association between the Department of Industrial and Vocational Education at the University of Alberta and practicing Industrial Arts teachers. Existing avenues of communication make it relatively easy to maintain contact with these people. Further, as a practicing Industrial Arts teacher, the researcher was better able to provide advice and assistance in using QUIZMASTER during the field test.

Lastly there was some additional incentive for the Industrial Arts teachers to participate in the field test because it was possible to make available to them a number of Industrial Arts Safety tests which were developed by J. F. D. Ilott at the University of Alberta. These tests were offered to the participating teachers and all the

participating teachers took advantage of the offer to some degree.

Collection of Data

The formal collection of data was done in two ways. First, all the participants in the study were asked to complete a questionnaire. The teachers and selected students were then interviewed to gain further information.

Collecting Data from the Students

There were two methods of collecting data from the participating students. The first involved the administration of a short questionnaire. The questionnaire was administered by the participating teachers to all of the classes that were involved in the study. The only students not completing a questionnaire were those students that were absent from class on the day the questionnaire was administered.

A structured interview was used as the second method of gathering data from the students. Only a few of the students were interviewed. The selection method was opportunistic. The students that were interviewed were selected because they were available at the time the researcher visited the school. Students who indicated an unwillingness to be interviewed were not interviewed.

The students were exposed to only a small part of QUIZMASTER. The information that could be obtained from the students was necessarily limited by that exposure. The

questionnaire and the interview items reflect this in that both are limited in length. The range of topics covered in these instruments is restricted to the act of actually using QUIZMASTER to complete a test.

Collecting Data from the Teachers

In contrast to the students, the teachers involved in the study saw all sides of the QUIZMASTER program. First, during the familiarization session prior to the field test starting, they had the opportunity to use the program as a student would use it. The authoring and record keeping functions were explored more extensively as the teachers used the program with their classes. In addition the participating teachers had the opportunity to observe student reaction when the students used the program as part of their Industrial Arts course of studies. This greater level of involvement is reflected in the greater breadth of questions asked in both the teacher questionnaire and the teacher interview schedule. These instruments can be found in Appendix 1.

Informal Sources of Data

Another source of useful information resulted from the activities involved in supervising the study. Some teachers encountered procedural problems which caused the loss of program data. Although the immediate solution for these problems was to suggest alternate methods of using the

program, the long term solution resulted in changes to the program code.

Chapter 4: Analysis and Interpretation of Data

Introduction

There were two distinct groups of people involved in the field test of QUIZMASTER. The largest group was made up of the students who used the program as part of their studies. Because this group is relatively large it is possible to come to some tentative conclusions regarding the students' perceptions of QUIZMASTER within the context of the Industrial Arts programs. The second group, made up of the teachers who participated in the field test, was quite small.

Because the group of teachers was so small, no reliable conclusions can be drawn from the data collected. This does not mean that the data collected is valueless. As Contract Research Corporation (1975) point out, in relation to field testing curriculum products:

> A simple unelaborated conclusion that no effects could be measured, however, would rob the potential users of important contextual data, including the insights of the actors involved. While intuition is insufficient basis for decision-making about a curriculum, it is important information for directing future exploratory efforts. (p. 60)

It is in this spirit that the data collected from the teachers is presented.

Referencing the Data

For reference, a summary of the data gathered in formal data gathering process is contained in Appendix 3. In the discussions which follow, number references to the Likert scale items are the numbers used to identify the statements in the data summaries in Appendix 3. As the data is examined, the relevant statements are reproduced in tables in the body of the text. Statement numbers starting with a 1 are taken from the student questionnaire and statement numbers starting with a 2 are taken from the teacher questionnaire. All other item number references are to the appropriate questionnaire item or interview schedule.

The process of calculating means for the Likert scale items meant that a value had to be assigned to each of the response categories. For the purposes of the calculations the "strongly disagree" category was assigned a value of one and the "strongly agree" category was assigned a value of five. The categories in between were assigned appropriate, intermediate values.

Examining the Data

The research questions posed by this study are quite specific. The answers to the individual research questions are important for two reasons. First they provide the information necessary to make informed judgments about improvements to be made to QUIZMASTER. Secondly, when taken

as a group, the answers provide clues to the answers to the broader questions that were detailed in the Statement of the Problem in Chapter 1. Those questions are: 1) Is QUIZMASTER's computer code correct and complete? 2) Is QUIZMASTER'S format and content acceptable? 3) Is QUIZMASTER acceptable to teachers and students?

The data collected in the field test is examined in two ways. First the individual questions are evaluated within the context of the way the research questions were grouped. The data is then examined in the context of the three broad questions outlined above.

Answering the Research Questions

The grouping of the research questions used in Chapter 1 provides a convenient way of examining the data. Each group of questions pays attention to particular aspects of the program's qualities. A three part format is used for examining the data. First there is a statement detailing the qualities of the program to be examined. Next follows a summary of the anticipated results based on certain, stated assumptions. Finally the data is examined to see if, in fact, the data asserts or denies the assumptions.

Computer Program code

This group of questions attempts to discover if any syntax or logic errors exist in the program code. While it is unlikely that a computer program the size of QUIZMASTER

will ever be completely free of code errors it should be possible to bring the number and type of errors within reasonable limits.

Anticipated Results

If the assumption is made that there are errors in the computer program code then a particular pattern of responses can be expected. First, mean values less than 5.00 should be expected for items 2.02 and 2.04 on the teacher questionnaire Likert scale data. The more errors that there are in the program, the lower these values should be. Specific complaints should show up in the responses to item 5 on the teacher questionnaire, items 1, 2, 9, and 10 in the teacher interview, and items 1 and 3 in the student interview.

Likert Scale Results

Examining the results of items 2.02 and 2.04 on the teacher questionnaire reveal means of 4.00 and 4.60 respectively. The statements and related data are detailed in Table 4-1. These values give an indication of the magnitude of the problem which existed during the field test. The relatively high values are important to note because, they imply that while were some errors in the program code, the bulk of the program, or at least that part of the program used by individual teachers, worked correctly and smoothly.

Table 4-1

Computer Program Code: Questions and Data Summary

(2.02) The program operated smoothly without producing error messages or hanging.

(2.04) The program responded correctly to commands.

Item Summary

Item #	SD	D	N	A	SA	Item N	Mode	Mean
2.02	0	0	1	3	1	5	4	4.00
2.04	0	0	0	2	3	5	5	4.60

Identifying The Errors

A study of the responses given to the relevant questions revealed several different program code errors. Item 5 on the teacher questionnaire reveals that one problem with the computer code was a failure to save student marks. This concern is echoed in the responses to questions 1, 2 and 9 of the teacher interview and question 3 of the student interview. The data indicates that this is not a very large problem from the students' point of view. Only two of the 64 students responding to question 3 of the student interview identified this as a problem. None of the 64 students responding to question 1 of the interview identified this as a problem. The small numbers notwithstanding, it is important to point out that any time

a computer program fails to record data it is a serious fault and requires remedial action.

Question 1 of the teacher interview identified a number of other errors with the program code. Without going into the technical details, the identified problems were:

1. Failure to read data correctly from the disk under certain circumstances. This resulted in an interruption in the program's operation and subsequent loss of data.
2. Failure to write data to the disk under certain circumstances.
3. Inadequate error trapping when the program was forced to deal with write protected disks and disk files.
4. The test generator would print incomplete test items under certain circumstances.

All of the identified errors, in one way or another, resulted in the loss of data from the program. In spite of the seriousness of the errors it was possible to avoid all of them by following certain, prescribed operating procedures. Informing the participants of the field test of the correct procedures made it possible to complete the field test without making modifications to the program.

There were a number of complaints about the slow speed of the program. These complaints showed up in question 9 of the teacher interview and questions 3 and 4 of the student questionnaire. This problem was rated second in the list of things students liked least about taking a test. It stems

from an early decision to make QUIZMASTER downward compatible with the older Apple II+ computers which were limited to 48 Kilobytes of memory. This requirement imposed certain restrictions on QUIZMASTER. First, it was necessary to use an older and slower disk operating system. Next it was necessary to use a filing system that used the disk as an extension of active memory. As a result the program does have some delays in operation. There is very little that can be done about the problem without changing the program specifications.

Computer Program Format and Content

The research questions, asked in this section, attempt to deal with the level of usability of the computer program in three broad areas; 1) ease of use, 2) the test format, and 3) the program completeness.

Anticipated Results: Ease of Use

There are three research questions associated with the ease of use. 1) Is the computer program user friendly? 2) Are the menu functions clear? 3) Is the screen format acceptable?

The means of the items listed in Table 4-2, taken together, should reflect the general level of difficulty in using the program. Each item references a different aspect of the problem. If the program is easy to use then the following pattern of responses would be expected. The means for items 2.01 and 2.20 should be at or near 1.00. Greater

levels of difficulty in using the program should be reflected in higher means for these items. Similarly the means for items 1.01, 1.02, 2.03, 2.05, 2.06 and 2.08 should be at or near 5.00. The more difficult the program is to use, the lower these means should be.

Problems of a specific nature should be found in the teacher questionnaire, questions 5 and 6, the teacher interview, questions 1, 2, 3, 4, and 9, the student questionnaire, questions 2, 3, and 4, and the student interview, questions 1 and 7.

Likert Scale Results: Ease of Use

From the student's perspective there are some problems associated with operating the program. Means of 3.73 and 4.22 on items 1.01 and 1.02 respectively indicate that the program is relatively easy to operate. Of the 244 students who completed the questionnaire, 142, or 58.2%, agreed or strongly agreed that the program could be operated without help from their teacher. There were 193 students, or 79.1%, who felt that the instructions on the screen were clear.

Table 4-2

Likert Data Summary: Ease of Use

(1.01) Once I had the computer disks I could run the program without much help from my teacher.

(1.02) The instructions on the screen were easy to understand.

(2.01) The program was difficult to use.

Table 4-2
Likert Data Summary: Ease of Use (continued)

(2.03) The instructions on the screen were clear.

(2.05) The menus were easy to understand.

(2.06) The screen layout was easy to read.

(2.08) My students found the program easy to use.

(2.20) I found the terminology used in the program (Unit, Topic, etc.) to be confusing.

Item Summary

Item #	SD	D	N	A	SA	Item N	Mode	Mean
1.01	14	21	59	62	80	236	5	3.73
1.02	7	14	29	61	132	243	5	4.22
2.01	3	0	1	0	0	4	1	1.50
2.03	0	0	1	2	2	5	5	4.19
2.05	0	0	2	2	1	5	4	3.80
2.06	0	0	0	1	4	5	5	4.80
2.08	0	0	0	2	3	5	5	4.60
2.20	1	3	1	0	0	5	2	2.00

Note. Item numbers starting with 1 are from the student's questionnaire and item numbers starting with 2 are from the teacher's questionnaire.

The results from the teacher questionnaire are not as reliable because of the smaller numbers. It is still interesting to note that the results indicate that teachers

participating in the field test found the program relatively easy to use. Item 2.01, which deals directly with the issue, has a mean of 1.50 indicating that the teachers found the program easy to use. Item 2.20 asks about the terminology used by the program. It shows a mean of 2.00 indicating that the terminology used by QUIZMASTER is relatively easy to understand. The screen layout appeared to be acceptable to the teachers, indicated by a mean of 4.80 for item 2.06. Items 2.03 and 2.05 had means of 4.19 and 3.80 respectively indicating that the messages displayed by the program in its menus and instruction statements were also easy to understand. None of the teachers rated the program as difficult to use.

Identifying the Problems: Ease of Use

From the students' point of view there is no clear statement that QUIZMASTER is easy to use. A Majority, or about 58%, did indicate that they found the program easy to use but this leaves over 40% of the students who presumably did encounter problems. When studying the anecdotal responses, many students appear to be satisfied that the program was easy to use. Question 2 on the student questionnaire asked "What did you like best about taking a test with a computer?". The most frequently occurring response, with a frequency of 68 out of 270, was that it was easy to use. It was not clear whether the majority of these students were referring to the program or the test

items included in the Unit. Similar results are found in the free responses to the student interview questions. When asked to identify any problems that they had with the program 34 out of 64 respondents indicated that there were no problems and when asked for general comments 13 out of 23 responded that they liked the program or found it easy to use. In both cases these were the most frequently occurring responses.

There were a few problems identified by the students. One respondent reported that the instructions on the screen were confusing, and three individuals reported having difficulty with the password. There is also evidence to indicate that some students, because of forgotten passwords, had difficulty with the security entry system. Another source of problems encountered by some students was the program's failure to store marks.

The teachers, who used more of the program, reported a greater number of problems. Even so the general tone was positive. Among the responses were comments to the effect that the program was straightforward and easy to use. One teacher pointed out that while it took some time and effort to learn the terminology used by the program, the benefit was that the structure helped to organize the thinking process when constructing a test.

Some of the anecdotal responses pointed to specific problems. The largest number of these related to the

complexity of the program. The comments ranged from requests for flow charts showing the program's organization to complaints that the menu options did not adequately describe the various program functions. There does appear to be a need for a simpler, more meaningful menu structure within the program.

The remaining comments relative to the ease of use of the program tended to point to specific problems such as particular messages displayed by the program or they were very general comments such as a request to make the program more user friendly

Anticipated Results: Test Format

There is only one research question associated with the test format. It asks; Is the test format acceptable? The more important aspects of the test format are identified in the relevant Likert scale statements listed in Table 4-3. A Unit, or Unit test, is made up of a number of basic components including questions, feedback statements, resource lists, and special information statements. These components can be assembled in a number of different ways, then presented to the student as a test. The result is that the same test can have a number of different appearances. The purpose of evaluating the test format is to determine if the individual components of the test are acceptable rather than looking at the various ways that the components can be assembled.

Of the items listed in Table 4-3, items 2.11 and 2.12 would be expected to have means at or near 1.00 if there is a high level of satisfaction with the statements. The remaining statements should have means at or near 5.00 if a high level of satisfaction is to be indicated. Clues to specific strengths or weakness may be found in the answers to questions 3, 4, 7, 8 and 9 of the teacher questionnaire, questions 7 and 8 of the teacher interview, questions 2 and 4 of the student questionnaire, and questions 2 and 5 of the student interview.

Likert Scale Results: Test Format

The components of a Unit that the student sees are predetermined by the teacher. For this reason the students were asked to respond to a more general statement indicating whether or not they liked the extra information given during the test review. Only 117 students, or 47.9%, either agreed or strongly agreed that they liked the extra information given during the test review. To balance this, only 35 disagreed or strongly disagreed with the statement. The remaining 67 students either had no opinion or declined to answer the question. At first glance the 47.9% figure may appear to be quite low but it must be pointed out that not all the students were given an opportunity to complete the test review. For this item it may be better to compare the number who agreed or strongly agreed with those who disagreed or strongly disagreed. On this basis there were

152 students who made a committed response. Of this number 117, or 77%, liked the information in the test review.

The teachers generally had a positive view about the various components that go into making up a QUIZMASTER Unit. With the exception of item 2.11, which deals with the number of mastery feedback statements used, all of the items have means which indicate a reasonably high level of teacher satisfaction.

Table 4-3

Likert Data Summary: Test Format

(1.04) I liked the extra information given during the test review.

(2.10) I liked the way question feedback statements were used.

(2.11) I used very few question mastery feedback statements.

(2.12) I used very few question non-mastery feedback statements.

(2.13) I used unit level mastery feedback statements.

(2.14) I used unit level non-mastery feedback statements.

(2.18) Four questions per topic is adequate.

(2.21) The ability to include resource lists with a test is useful.

Table 4-3

Likert Data Summary: Test Format (continued)

Item Summary								
Item #	SD	D	N	A	SA	Item N	Mode	Mean
1.04	37	18	65	51	66	237	5	3.38
2.10	0	0	0	2	3	5	5	4.60
2.11	1	1	0	3	0	5	4	3.00
2.12	3	1	1	0	0	5	1	1.60
2.13	0	1	0	2	2	5	5	4.00
2.14	0	1	0	2	2	5	5	4.00
2.18	0	0	0	2	3	5	5	4.60
2.21	0	0	0	2	3	5	5	4.60

Note. Item numbers starting with 1 are from the student's questionnaire and item numbers starting with 2 are from the teacher's questionnaire.

Identifying Test Format Problems

Perusing the anecdotal responses gleaned from the student questionnaire and the student interviews provides some information as to what the students liked or did not like. On the positive side, some students liked the types of questions used and liked having immediate feedback. These were the fifth and sixth most frequently occurring responses to question 2 on the student questionnaire. In

response to question 2 of the student interview, a number of students liked the idea that the program provided feedback and gave them their marks right away. This was the third ranked comment in response to this question.

During the student interview the students were asked specifically if they found the feedback screens and resource lists useful. Of the sixty-two students responding to this question twenty-six did not recall seeing any feedback screens. All but one of the remaining thirty six students found that these screens helped in one way or another.

The complaints that the students registered applied more to the way the various components of the QUIZMASTER unit were used rather than the focusing on the components themselves. There were conflicting comments about the test review. Some students complained about being forced to complete the test review while others felt that the test review was a good experience. A third group, presumably made up of those students who were denied a test review, registered the complaint that there was no test review. The incidence of these comments was small with the number of students being limited to only two or three for each comment.

In general the teachers who participated in the field test appeared to be satisfied with the components that make up the units. During the interview the teachers were asked

what features were unnecessary and what features should be added. These questions were intended to catch shortcomings. In general, however, it would be reasonable to expect comments relative to the test format to be included in the responses. There were no program features identified as unnecessary and there were no requests for new components to be added to the QUIZMASTER Unit. The relevant comments that were made all requested greater control over how the various components could be assembled to make a test.

Anticipated Results: Program Completeness

The research questions in this section attempt to discover if there are functions which could be added to QUIZMASTER in order to make it more useful or if there are any functions already incorporated which are not likely to be used. There are three relevant research questions: 1) Have any important functions been omitted? 2) Are there any superfluous functions? 3) Are premade tests desirable?

Items 2.22 and 2.23 on the teacher questionnaire Likert scale attempt to determine if the grade book and the item analysis modules, built into QUIZMASTER, are useful functions. If the teachers perceive these as useful items then the mean of item 2.22 should be at or near 5.00 and the mean of item 2.23 should be at or near 1.00. Items 2.15 and 2.16 attempt to determine whether QUIZMASTER might be of more use to the classroom teacher if it were supplied with a number of Units relevant to the teacher's subject

area. Since the teachers participating in the field test were supplied a number of prepared units and were asked to create an unit of their own, they have some basis for comparison. If ready made tests are desirable, the means for items 2.15 and 2.16 should be at or near five. The anecdotal responses in the questionnaires and interviews were relied upon to provide clues to any necessary changes in program features. The questions most likely to contain relevant responses are questions 4, 6, and 10 on the teacher questionnaire and questions 4, 7, and 8 of the teacher interview.

Likert Scale Results: Program Completeness

The relevant Likert scale items and the associated data are shown in Table 4-4.

There is no clear consensus, among the teachers who participated in the field test, whether ready made tests are preferable. Means of 3.00 and 2.80 for items 2.15 and 2.16 respectively indicate that the teachers are divided on this issue. In order to arrive at a meaningful answer for this question a much larger number of teachers will need to be surveyed.

The grade book does seem to be a useful feature in the program. Item 2.20 has a mean of 4.40 indicating that the participating teachers saw value in it.

Table 4-4

Likert Data Summary: Program Completeness

(2.15) I would prefer a ready made test bank.

(2.16) It takes too much time to create my own test bank.

(2.22) The grade book format was useful.

(2.23) I would not use the item analysis data collected by the program.

Item Summary

Item #	SD	D	N	A	SA	Item N	Mode	Mean
2.15	2	0	0	2	1	5	4	3.00
2.16	1	1	1	2	0	5	4	2.80
2.22	0	0	0	3	2	5	4	4.40
2.23	2	1	1	1	0	5	1	2.20

A mean of 2.20 for item 2.23 and a response pattern that ranges from strongly disagree up to agree indicates that the teachers were divided on the usefulness of the item analysis module.

Identifying the Shortcomings

None of the teachers suggested any new functions which should be added to QUIZMASTER or any features which could be classed as unnecessary. One teacher, however, did advocate adding a routine to the Student Disk which would

print a completion certificate for the student on the successful completion of a test.

There were a number of valuable comments and a number of suggestions on how to improve the existing features. One of the teachers pointed out that the time required to author a test was one of the least liked features of the program. This comment is consistent with the Likert scale results. Some comments on the grade book suggested that the grade book use raw scores rather than percentage scores, that the grade book be expanded to handle more students, and that it allow the entry of other grades generated outside of QUIZMASTER. There were very few comments about the item analysis module but one teacher did like the capability of obtaining feedback on question quality.

Computer Program Acceptability

The last group of research questions deal with specific aspects of the acceptability of the computer program. Three questions were asked: 1) Will students accept this form of testing? 2) Do teachers see the computer program as a time saving device? 3) Is there sufficient value in the computer program for teachers to create their own test banks?

Anticipated Results

There are a number of Likert scale items on both the teacher and student questionnaires that attempt to answer the research questions in this section. These items are

listed in Table 4-5. Items 1.03, 1.05, and 1.06 deal with the students' perception of QUIZMASTER. If the program represents an acceptable form of testing then the mean for item 1.05 should be at or near 1.00 and the means for items 1.03 and 1.06 should be at or near 5.00. The teachers' view of QUIZMASTER is, naturally, much broader than that of the students. The remaining items in Table 4-5 try to sample some of the more important concerns that a teacher might have. If, from the teachers' point of view, the computer program represents an acceptable form of testing and, at the same time, meets the needs of the teacher in the classroom, then the following response patterns can be expected. Items 2.07 and 2.16 should have means at or near 1.00 while the remaining items should have means at or near 5.00.

Likert Scale Results

The results from the items seem to indicate that the students are willing to accept the form of testing presented by QUIZMASTER. Of the 244 students responding to the questionnaire, 198 or 81.5% agreed or strongly agreed with the statement that they liked using the computer to take a test. In a similar vein, 189 students, or 77.5%, disagreed or strongly disagreed with the statement that they preferred writing tests with a paper and pencil. When it came to assessing how much the program aided as a learning tool, the results were quite scattered. 45.5% of

the students felt that it did help while 20.5% felt that it did not. The mean for item 1.06 was 3.31. The teachers, who might be in a better position to assess this concept, felt that using QUIZMASTER did help improve learning. The mean for item 2.24 is 4.80.

Item 2.19 addresses the issue of whether the teachers see the computer program as a time saving device. The mean in response to this item is 4.80. All of the teachers either agreed or strongly agreed that the computer program would save them time in the classroom.

Table 4-5

Likert Data Summary: Program Acceptability

(1.03) I liked using the computer to take a test.

(1.05) I prefer writing tests with a paper and pencil.

(1.06) The testing program helped me learn the subject better.

(2.07) The program was not very useful.

(2.09) I would use this program in my classroom/laboratory.

(2.16) It takes too much time to create my own test bank.

(2.17) The program fills a need in my classroom/laboratory.

(2.19) This program will save me time in the classroom.

(2.24) This form of testing helps improve learning.

Table 4-5

Likert Data Summary: Program Acceptability (continued)

Item Summary								
Item #	SD	D	N	A	SA	Item N	Mode	Mean
1.03	19	9	17	69	129	243	5	4.15
1.05	149	40	25	12	17	243	1	1.79
1.06	28	22	81	67	44	242	4	3.31
2.07	4	1	0	0	0	5	1	1.20
2.09	0	0	0	0	5	5	5	5.00
2.16	1	1	1	2	0	5	4	2.80
2.17	0	0	0	1	4	5	5	4.80
2.19	0	0	0	1	4	5	5	4.80
2.24	0	0	0	1	4	5	5	4.80

Note. Item numbers starting with 1 are from the student's questionnaire and item numbers starting with 2 are from the teacher's questionnaire.

The third research question in this section touches on two different concepts. First it attempts to assess the value of the computer program to teachers. It then asks specifically if there is sufficient value to encourage the teachers to do the front end work of creating a number of Units for use in the classroom. It becomes apparent, from examining the results of items 2.07, 2.09 and 2.17, that

the teachers do see value in using QUIZMASTER. In terms of preparation time, however, there is no clear consensus. Item 2.16, which deals with the time required to author a Unit, did not yield a clear answer. It has a mean of 2.80, which is inconclusive, and the responses range from strongly disagree to agree.

Discussion of the Results

In the statement of the problem in chapter 1, three major questions were posed. The first of these asked if the computer code was correct and complete. No matter how well designed a computer program might be, it cannot perform its function if the operating code is not correct. The field test did identify a number of errors in the operating code. The majority of these were disk access problems that occurred when data was either stored or retrieved from disk storage. The problems did not interrupt the field test, however, since it was possible to devise operating procedures which allowed for the safe storage and retrieval of data.

The second question asked if QUIZMASTER's format and content were acceptable. The question cannot be answered with a simple yes or no. Some general answers, however, can be arrived at by summarizing the results of the preceding discussions.

The primary purpose of QUIZMASTER is to provide a capability for interactive testing. If the program is to be

successful the test format must be acceptable to both the students and the teachers. If the format was not acceptable the rest of the computer program would become irrelevant.

In general both the teachers and the students found the test format acceptable. A majority of the students liked the program features which presented additional information during the test. The teachers as well had a positive view about being able to provide additional information for their students. There is evidence to show that the various components of a QUIZMASTER Unit were not used uniformly. Question Mastery Feedback statements, for example, tended to be used somewhat less than the other Unit components. There was also evidence to show that the teachers participating in the field test preferred multiple choice questions over short answer questions. Despite this there was no indication that any of the components of the QUIZMASTER Unit were considered unnecessary by the participating teachers. Similarly there were no requests for any new components to be added.

For any computer program to be received favourably it must be relatively easy to use. In this particular area QUIZMASTER falls a little short. Only 58% of the students reported that they could operate the program without much help from their teacher. The problems encountered by the students were serious enough to be reflected in some of the comments made by the teachers.

Further comments made by the teachers indicated that some of the menu items were not clear. Several of the teachers felt that, because of the complexity of the program, it was easy to get lost. There were several suggestions on how to improve this aspect of the program.

If QUIZMASTER is to be a useful tool for the teacher it must perform all, or nearly all, of the expected functions. The teachers participating in the field test did not report any major shortcomings in QUIZMASTER. The suggestions that were made tended to ask for expanded capabilities of the existing functions. There was, however, some doubt about whether QUIZMASTER should be supplied with ready made Units and whether the item analysis module would prove to be a useful program function. In the case of the latter, the program was not used long enough to gather sufficient data to properly assess the function.

The third question posed by the statement of the problem asked whether QUIZMASTER was acceptable to teachers and students. From an overall perspective the teachers participating in the field test found QUIZMASTER to be generally acceptable. All of them felt that the program did aid the learning process and filled a need in their classrooms. In addition the program did save the teachers some time. They indicated that they would use the program in their classrooms.

From the students' perspective there was a clear indication that the majority liked using the computer to take a test and that they appreciated knowing their test results immediately.

Although there were a number of specific shortcomings identified by the field test it appears that QUIZMASTER was acceptable program to both the teachers and the students who participated in the field test.

Chapter 5: Conclusions and Recommendations

Purpose of the Study

The purpose of this study was to evaluate the acceptability of the computer assisted testing program, QUIZMASTER. The problem was defined by three questions: Is QUIZMASTER's computer code correct and complete? Are QUIZMASTER'S format and content acceptable? Is QUIZMASTER acceptable to teachers and students? To find the answers to these questions it was necessary to formulate and answer a number of research questions. The research questions were organized into three corresponding categories: computer program code, computer program format and content, and computer program acceptability. Answers to the research questions were sought by conducting a field test in the Industrial Arts laboratories of four junior high schools and two senior high schools. Five teachers and 244 students participated.

The field test was a part of a longer sequence of events used to evaluate QUIZMASTER during its developmental period. The sequence began with drafting the program specifications and progressed through a period of pilot testing to reach the field test phase.

The formal data gathering process, for the field test, had two parts. All of the teachers and students participating in the field test were asked to complete a questionnaire containing both likert scale items and

questions calling for an anecdotal response. Interviews were then conducted with all of the participating teachers and with selected students. The data that was collected is summarized in Appendix 3.

Summary of the Results

It is important to keep in mind that the field test was limited in its size and scope. Some limited conclusions, based on the data gathered from the 244 students who participated in the field test, are presented in this chapter. With only five teachers responding, it is not possible to extrapolate beyond those participating in the field test. This does not mean that the data obtained from them was not useful; the insights of the teachers provided valuable clues on how to proceed in the development of QUIZMASTER.

Computer Program Code

The first question posed in the statement of the problem dealt with the program code. No matter how well designed a computer program might be, it cannot perform its function if the operating code is not correct. Further, errors in the code are likely to discourage potential users from using the program thus rendering it an unacceptable product. There were two research questions related to the computer program code:

1. Does the computer program cease operation at unexpected times?

2. Does the computer program respond to commands and data input in the expected manner?

The field test did identify several errors in the operating code. The majority of these were disk access problems that occurred when data was either stored or retrieved from disk storage. The problems did not interrupt the field test since it was possible to devise a procedure which allowed for the safe storage and retrieval of data. Although experience is limited by the low number of teachers participating, the evidence indicated that the bulk of the program operated correctly, and smoothly with only a few errors in the program code.

Computer Program Format and Content

The second question in the statement of the problem asked if QUIZMASTER's format and content were acceptable. The research questions relevant to this aspect of the study are:

1. Is the computer program user friendly?
2. Are the menu functions clear?
3. Is the screen format acceptable?
4. Have any important functions been omitted?
5. Are there any superfluous functions?
6. Is the test format acceptable?
7. Are premade tests desirable?

To simplify the discussion, the research questions in this group were divided into three subgroups. The first

subgroup focused on questions 1, 2, and 3 which dealt with different aspects of the ease of program operation. The second subgroup centred around question 6 which dealt with the acceptability of the test format. The remaining questions were related to the issue of program completeness.

Ease of Use. For any computer program to be received favourably it must be relatively easy to use. The results of the study showed that QUIZMASTER was easy to use. It appears, however, that it takes some time to learn to use the program. A large minority of students reported that they needed help from their teachers to operate the program. The problems encountered by the students were serious enough to be reflected in comments made by the teachers. Further comments made by the teachers indicated that there were some difficulties with the program's terminology, the clarity of some statements displayed on the screen, and adjusting to the complexity of the program. The problems that were encountered were related to individual program functions and operating procedures and not with the fundamental organization of the program.

Test Format. In general both the teachers and the students found the test format acceptable. A majority of the students liked the program features which presented additional information during the test. The teachers as well had a positive view about being able to provide

additional information for their students. There is evidence to show that the various components of a QUIZMASTER Unit were not used uniformly by the teachers participating in the field test. Question Mastery Feedback statements, for example, tended to be used somewhat less than the other Unit components and multiple choice questions were preferred over short answer questions. There was no indication that any of the components were considered unnecessary. Similarly there were no requests for any new components to be added to the QUIZMASTER Unit.

Program Completeness. The teachers participating in the field test did not report any major shortcomings in QUIZMASTER. The suggestions that were made tended to ask for expanded capabilities of the existing functions such as greater capacity in the program's grade book and increased control over testing conditions. There was, however, some doubt about whether QUIZMASTER should be supplied with ready made Units and whether the item analysis module would prove to be a useful program function. In the latter case the program was not used long enough to accumulate enough data to properly assess the item analysis function.

Computer Program Acceptability

The point of conducting the field test was to determine the acceptability of QUIZMASTER to teachers and students. In order to round out the information collected

it was necessary to ask a few direct questions relating to specific aspects of the acceptability question.

The majority of students seem ready to accept computer based testing into the classroom. They indicated that they liked using the computer to take a test and preferred using the computer over writing paper and pencil tests. It should be remembered, however, that QUIZMASTER was only used for a period of about 3 months and the novelty of using the computer to take a test may be interfering with the results.

The teachers participating in the field test found QUIZMASTER to be acceptable and indicated a willingness to use the program in their classrooms. All of them felt that the program aided the learning process, filled a need in their classrooms and saved time in the classroom.

Modifying the Program

The results of the field test indicated a number of areas where QUIZMASTER could be improved. It was decided that the some of the suggested changes should be made before the program was used again. Some of the more important changes that were made included the manner in which the QUIZMASTER Teacher Disk sets up and accesses the QUIZMASTER Units. The program menus were altered so that it would be easier to access the different functions built into QUIZMASTER. Major changes were made to the program function which allows the teacher to set the testing

conditions for the student disk. The editing capabilities of this function were expanded as were the number of parameters under teacher control. Although there is evidence that some students encountered difficulties in using the computer program, the cause of the problem was unclear. As a result, no major changes were made to the operating procedures on the student disk.

The teachers made many valuable suggestions on how to improve QUIZMASTER. Most of those suggestions have been incorporated as changes to the computer program. Limitations imposed by the computer hardware make it impossible to incorporate all of the suggested changes without making important changes to the program specifications.

Fixing The Code Errors

It was important to pay attention to the errors that were identified in the computer program code. Those code errors that had solutions were corrected, however, it was not possible to find solutions for all of the identified problems within the existing program specifications.

One perplexing problem was the program's occasional failure to store student marks. An investigation of this complaint did not reveal any problems in the program code. There does not appear to be a satisfactory solution to this problem short of changing the program specifications to

take advantage of recent improvements in the computer hardware.

There was a complaint that the program took a long time to retrieve question data from the disk which also went unresolved. The primary cause of this problem was the file structure used for the QUIZMASTER question files. That structure was adopted because the original program specifications called for the program to operate on an Apple II series computer with only 48 Kilobytes of memory. The solution to this problem is to change the program specifications so that it is possible to take advantage of a newer and faster computer hardware

The remaining identified errors in the program code were eliminated. There were three basic types of errors; 1) failure to process the disk operating system commands properly, 2) errors in reading information back from the data files, and 3) inadequate error trapping when accessing data files on the program disks.

Conclusions

In general, QUIZMASTER seems to be an acceptable program to the teachers who participated in the field test. The program filled a need in their classrooms, both administratively and pedagogically. The results indicated that the program would save the teacher time in the classroom and should assist the student in learning the subject matter. The teachers did point out a number of

shortcomings in the program. Most of these were either recommendations for procedural changes or requests for increased capacities in the various program functions. No major deficiencies were identified.

Because of the small number of teachers involved it is important not to extrapolate the results beyond the individuals concerned. The problem is further complicated by the nature of the Industrial Arts program. The program's curriculum places the Industrial Arts teachers in a position where they must individualize instruction. Even if the numbers were larger the results could not be generalized to the more traditional academic classrooms where the approach to instruction is more group oriented.

There is evidence to indicate that students are prepared to accept computer assisted testing. The results showed that the majority of students liked using the computer to take a test and that they preferred the computer tests over paper and pencil tests. Although the novelty of using a computer for classroom testing is undoubtedly an interfering factor, the majority of students favouring the system does seem large enough to support the conclusion.

The results of the study are positive enough to encourage further development of QUIZMASTER. The responses from the students indicate that they are prepared to accept computer assisted testing in the classroom.

In spite of the numbers involved, the data gathered from the teachers has served as a useful guide in re-structuring the program. Many of the problems that they identified only needed to be identified by a single individual to make the observation valid. Program code errors, for example, fall into this category. There were a number of observations which were subjective. These observations, too, have value. It should be remembered that the teachers participating in the field test, by virtue of their years of experience in the classroom, may be considered knowledgeable observers of test construction and administration. Their observations, therefore, were considered seriously.

Recommendations

The following recommendations are made for the continued development of QUIZMASTER.

1. The development of QUIZMASTER should continue.
2. The program specifications should be reviewed to allow the program to take advantage of recent advances in computer hardware in increased memory capacity. This implies that the computers of choice will be the enhanced Apple IIe, the Apple IIc, and the Apple IIgs.
3. The speed of operation of the program should be increased by changing the disk operating system and the file structures to take advantage of improvements in system hardware.

4. Where possible the data storage capability of the various program functions should be expanded to the limits imposed by the existing hardware.

The following recommendations are made for future testing of QUIZMASTER.

1. Once the changes are made to the program code QUIZMASTER should be again subjected to a field test to determine its suitability for use in a variety of classroom situations.
2. During the field test data should be gathered to determine the relative importance of the various test components under a variety of testing situations.

Bibliography

Alberta Education (Undated). Junior high school Grades 7-8-9: Industrial Education Curriculum. Edmonton, Alberta: Alberta Education.

Alessi, S. M. & Trollip S. R. (1985). Computer-Based Instruction, Methods and Development. Englewood Cliffs, N.J.: Prentice-Hall.

Anderson, John (1985). Software evaluation. In K. Alexander & D. Blanchard (Eds.), Educational Software: A creator's handbook (pp. 197-227). Loughborough, England: Tecmedia Limited.

Bloom, B. S., Hastings, J. T, & Madaus, G. F. (1971). Handbook on Formative and Summative Evaluation of Student Learning. New York: McGraw-Hill

Cartwright, G. F. & Derevensky (1976). An attitudinal study of computer-assisted testing as a learning method. Psychology in Schools. 13(3)

Clearinghouse of Alberta Education (1984?). Clearinghouse evaluator's guide for microcomputer-based courseware. Edmonton, Alberta: Alberta Education

Contract Research Corporation (1975). Guidelines for pilot and field testing of curriculum products. (For the Bureau of Occupational and Adult Education, U. S. Office of Education). Belmont Mass.: Contract Research Corporation.

Dunkleberger, G. & Heikkinen, H. (1983). Computer-made tests set mastery learning free. Curriculum Review, 22(5).

Flast, L. & Flast, R. (1986). Appleworks Applications. Berkeley, California: Osborne McGraw-Hill.

Gronlund, N. E. (1981). Measurement and evaluation in teaching (4th ed.). New York: Macmillian.

Gronlund, N. E. (1982). Constructing Achievement Tests. Englewood Cliffs, N.J.: Prentice-Hall.

Hambleton, R. K. (1984). Using microcomputers to develop tests. Educational Measurement. 3(2), p 10.

Hambleton, R. K., Anderson, G. E., & Murray, L. N. (1983, September). Applications of microcomputers to classroom testing. New Directions for Testing and Measurement, (19).

Harris, N. D. C., Bell, C. D., & Carter, J. E. H. (1981). Signposts for Evaluating: A resource pack. Leicester, England: H. Cave & Co. Ltd.

Hsu, T. & Nitko, A. J. (1983). Microcomputer testing software teachers can use. Educational Measurement: Issues and Practice. 2(4), p. 15.

Hsu, T. & Sadock, S. F. (1985). Computer Assisted Test Construction: The State of the Art. Princeton, New Jersey: ERIC Clearinghouse on Tests, Measurement, and Evaluation.

Jelden, D. L. (1982). Computer Generated testing. AEDS Monitor, 20(7)

Kearsley, G. (1983). Computer Based Training A Guide to Selection and Implementation. Reading, Massachusetts: Addison-Wesley.

Lathrop, A. & Goodson, B. (1983). Courseware in the classroom. Menlo Park, CA: Addison-Wesley.

Layman, J. & Hall, W. (1986). Teaching about the electricity supply industry-the development and evaluation of a computer based teaching package. Computers and Education, 10(1).

Miller, L. & Burnett, J. D. (1986). Theoretical considerations in selecting language arts software. Computers and Education, 10(1).

Millman, J. (1984). Using microcomputers to administer tests: An alternate point of view. Educational Measurement: Issues and Practice, 3(2), P. 20.

Millman, J. & Arter, J. A. (1984). Issues in item banking. Journal of Educational Measurement. 21(4).

Mizokawa, D. T. & Hamlin, M. D. (1984). Guidelines for computer-managed testing. Educational Technology, 24(12).

Nelson, L. R. (1984). Using microcomputers to assess achievement and Instruction. Educational Measurement: Issues and Practice, 3(2), p. 22.

Noble J. W. (1981). Computerized testing: a new system of evaluation. The Mathematics Teacher, 74(5)

Popham, W. James (1981). Modern Educational Measurement. Englewood Cliffs, N.J.: Prentice-Hall.

Rawitsch, D. (1983). Evaluating computer courseware: even old dogs need only a few new tricks. Social Education, 1147(5)

Roeber, E. D. (1984). Using Microcomputers to assess achievement and instruction. Educational Measurement: Issues and Practice, 3(2), p. 27.

Senn, P. R. (1983) Six checklists to prepare your classroom for technology. Social Education, 47(5)

Ward, W. C. (1984). Using microcomputers to administer tests. Educational Measurement: Issues and Practice, 3(2), P. 16.

Appendix 1

Field Test Questionnaires and Interview Schedules

THE UNIVERSITY OF ALBERTA
DEPARTMENT OF INDUSTRIAL AND VOCATIONAL EDUCATION

QUIZMASTER EVALUATION QUESTIONNAIRE

STUDENT

1. Please rate each of the following statements by circling the appropriate response. A rating of "1" means that you STRONGLY DISAGREE with the statement and a 5 means that you STRONGLY AGREE with it. Leave the item blank if you don't know.

	SD				SA
Once I had the computer discs I could run the program without much help from my teacher.	1	2	3	4	5
The instructions on the screen were easy to understand.	1	2	3	4	5
I liked using the computer to take a test.	1	2	3	4	5
I liked the extra information given during the test review.	1	2	3	4	5
I prefer writing tests with a paper and pencil.	1	2	3	4	5
The testing program helped me learn the subject better.	1	2	3	4	5

2. What did you like best about taking a test with a computer?

.....

3. What did you like least about taking a test with a computer?

.....

4. Please write any comments you may have about using this computer program.

.....

.....

THE UNIVERSITY OF ALBERTA
DEPARTMENT OF INDUSTRIAL AND VOCATIONAL EDUCATION

STUDENT INTERVIEW QUESTIONS

1. What kinds of things gave you problems with the program?
2. What did you like best about the program?
3. What did you like least about the program?
4. Did you do the test review?
5. Were the feedback screens and resource lists useful?
6. Would you like a last chance to go over the test before it is marked?
7. Do you have any comments or suggestions you would like to make?

THE UNIVERSITY OF ALBERTA
DEPARTMENT OF INDUSTRIAL AND VOCATIONAL EDUCATION

QUIZMASTER EVALUATION QUESTIONNAIRE

TEACHER

1. HARDWARE DESCRIPTION

For each of the computer types listed below please indicate the number of computers you have that are equipped with a single disc drive and the number of computers that are equipped with two disc drives.

	Number with single drive	Number with two drives
Apple II+		
Apple //e		
Apple //c		
Other (specify)		

How many printers do you have? _____

What type(s) of printers do you have? _____

2. Please rate each of the following statements by circling the appropriate response. A rating of "1" means that you STRONGLY DISAGREE with the statement and a 5 means that you STRONGLY AGREE with it.

	SD				SA
The program was difficult to use.	1	2	3	4	5
The program operated smoothly without producing error messages or hanging.	1	2	3	4	5
The instructions on the screen were clear.	1	2	3	4	5
The program responded correctly to commands.	1	2	3	4	5
The menus were easy to understand.	1	2	3	4	5
The screen layout was easy to read.	1	2	3	4	5
The program was not very useful.	1	2	3	4	5
My students found the program easy to use.	1	2	3	4	5
I would use this program in my classroom/laboratory.	1	2	3	4	5
I liked the way question feedback statements were used.	1	2	3	4	5
I used very few question mastery feedback statements.	1	2	3	4	5
I used very few question non-mastery feedback statements.	1	2	3	4	5
I used unit level mastery feedback statements.	1	2	3	4	5
I used unit level non-mastery feedback statements.	1	2	3	4	5
I would prefer a ready made test bank.	1	2	3	4	5
It takes too much time to create my own test bank.	1	2	3	4	5
The program fills a need in my classroom/laboratory.	1	2	3	4	5
Four questions per topic is adequate.	1	2	3	4	5
This program will save me time in the classroom.	1	2	3	4	5

I found the terminology used in the program (Unit, Topic, etc.) to be confusing.	1 2 3 4 5
The ability to include resource lists with a test is useful.	1 2 3 4 5
The grade book format was useful.	1 2 3 4 5
I would not use the item analysis data collected by the program.	1 2 3 4 5
This form of testing helps improve learning.	1 2 3 4 5

3. What one feature did you like most about the program?

.....

4. What one feature did you like least about the program?

.....

5. What was the most frequently occurring problem you had in running the program?

.....

.....

6. What two or three things should be changed to make the program more useful?

.....

7. What testing options beside feedback, mastery level, and access, would you like to see under teacher control?

.....

.....

8. What type of question did you use most?

.....

9. What type of question did you use least?

.....

10. Are there any functions that should be added? (If yes please describe)

.....

THE UNIVERSITY OF ALBERTA
DEPARTMENT OF INDUSTRIAL AND VOCATIONAL EDUCATION

TEACHER INTERVIEW QUESTIONS

1. What kinds of things gave you problems with the program?

2. Did your students have any problems with the program? If so what problems did they have?

3. What do you think are the greatest weaknesses of the program organization?

4. What did you like about the program organization?

5. Is the program format (Units, Topics, etc.) effective? If not what would you change?

6. How would you use this program in your classroom?

7. What features of the program are unnecessary?

8. What features would you like to see added?

9. Did you find any aspect of the program operation annoying?

10. Do you have any comments or suggestions that you would like to make?

Appendix 2

Pilot Test Questionnaire and Program Fault Report

TESTING PROGRAM QUESTIONNAIRE

1. Please rate each of the following statements by circling the appropriate response. A rating of "1" means STRONGLY DISAGREE with the statement and a 5 means that you STRONGLY AGREE with it.

	SD				SA
The program was difficult to use.	1	2	3	4	5
The instructions on the screen were clear.	1	2	3	4	5
The menu functions were confusing.	1	2	3	4	5
I had adequate preparation before using the program.	1	2	3	4	5
The program was not very useful.	1	2	3	4	5
My students found the program easy to use.	1	2	3	4	5
I would use this program in my classroom.	1	2	3	4	5
I liked the way feedback statements were used.	1	2	3	4	5
I would prefer a ready made test bank.	1	2	3	4	5
I would not take the time to create my own test bank.	1	2	3	4	5
Four questions per topic is adequate.	1	2	3	4	5
Using separate teacher and student discs is a problem.	1	2	3	4	5
This program would save me time in the classroom.	1	2	3	4	5
I found the terminology used in the program (Unit, topic, etc.) confusing.	1	2	3	4	5

2. What one feature did you like most about the program?

.....

3. What one feature did you like least about the program?

.....

4. What was the most frequently occurring problem you had in running the program?

.....

.....

5. What two or three things should be changed to make the program more useful?

.....

DEPARTMENT OF INDUSTRIAL AND VOCATIONAL EDUCATION

PROGRAM FAULT REPORT

NAME ..... DATE .....

Disc name (Check one)

[] Teacher disc
[] Student disc
[] Question disc

Version number .....

TYPE OF ERROR (Check one)

[] Fatal (Program stopped)
[] Fatal (Lost or changed data)
[] Nonfatal (Program caused confusion as how to proceed)

INDICATE THE NATURE OF THE PROBLEM

1. Give error message if any. Include line number if given.

   .....

2. Which disc was the program using at the time of failure. (If known)

   .....

3. Please give the name of the module or function in use at the time of failure.

   .....

4. Describe what you were doing at the time of failure.

   .....

5. Provide any other information that may be helpful in finding the program error.

   .....

Please return the completed report to Wally Latta.

Appendix 3: Summary of Data

Introduction

For the purposes of the field test the computer program, QUIZMASTER, was placed in six junior high schools and two senior high schools. Two of the junior high schools did not use the program and were, therefore, unable to provide any data. One of the teachers in the remaining junior high schools left his position in the school before any data could be obtained from him. He could not be located. Despite this it was possible to obtain data from the students in that school. Consequently data was obtained from six schools in all, four junior high schools and two senior high schools. Five teachers and 244 students participated in the field test.

The formal data collection process resulted in data in three different forms. All of the participants in the field study were asked to complete a questionnaire. The questionnaires contained two distinct types of questions. First the participants were asked to rate a series of statements on a five response Likert scale. The second form for the data obtained from the questionnaires was written comments in response to specific questions. The third form for the data was notes made by the researcher in interviewing the participating teachers and selected students.

For reference, the questionnaires completed by the students and teachers can be found in Appendix 1.

Teacher Questionnaire Data Summary

The questionnaire was completed by five teachers, three junior high school teachers and 2 senior high school teachers. Because of the small numbers the data has been combined. For convenience the questionnaire items are reproduced with each of the data summaries.

Hardware description

The teachers were asked to provide information on the type and configuration of the hardware they had available for use. This information was requested in anticipation of hardware incompatibilities embedded within QUIZMASTER. Since no incompatibility problems arose during the field test this information was not required. The question that the teachers were asked to respond to is listed below. The information is tabulated in Table A3-1.

Question: For each of the computer types listed below please indicate the number of computers you have that are equipped with a single disk drive and the number of computers that are equipped with two disk drives.

Table A3-1

Hardware Data

Equipment Type	Number with single drive	Number with two drives	Total
Apple II+ computer	2	1	3
Apple //e computer	5	3	8
Apple //c computer	0	1	1
other computer	0	0	0
Imagewriter printer			2
Imagewriter I printer			1
Imagewriter II printer			1
Roland 1111A printer			1

Likert Scale Data

Teachers were asked to respond to a number of statements by indicating their relative agreement on a five point rating scale. For ease of reference the statements are reproduced below. The item numbers associated with the statements did not appear on the questionnaire. They have been added for reference when interpreting the data summarized in Table A3-2.

(2.01) The program was difficult to use.

(2.02) The program operated smoothly without producing error messages or hanging.

(2.03) The instructions on the screen were clear.

(2.04) The program responded correctly to commands.

(2.05) The menus were easy to understand.

(2.06) The screen layout was easy to read.

(2.07) The program was not very useful.

(2.08) My students found the program easy to use.

(2.09) I would use this program in my classroom/laboratory.

(2.10) I liked the way question feedback statements were used.

(2.11) I used very few question mastery feedback statements.

(2.12) I used very few question non-mastery feedback statements.

(2.13) I used unit level mastery feedback statements.

(2.14) I used unit level non-mastery feedback statements.

(2.15) I would prefer a ready made test bank.

(2.16) It takes too much time to create my own test bank.

(2.17) The program fills a need in my classroom/laboratory.

(2.18) Four questions per topic is adequate.

(2.19) This program will save me time in the classroom.

(2.20) I found the terminology used in the program (Unit, Topic, etc.) to be confusing.

(2.21) The ability to include resource lists with a test is useful.

(2.22) The grade book format was useful.

(2.23) I would not use the item analysis data collected by the program.

(2.24) This form of testing helps improve learning.

Table A3-2

Likert Data Summary - Teacher

Item Summary								
Item #	SD	D	N	A	SA	Item N	Mode	Mean
2.01	3	0	1	0	0	4	1	1.50
2.02	0	0	1	3	1	5	4	4.00
2.03	0	0	1	2	2	5	5	4.19
2.04	0	0	0	2	3	5	5	4.60
2.05	0	0	2	2	1	5	4	3.80
2.06	0	0	0	1	4	5	5	4.80
2.07	4	1	0	0	0	5	1	1.20
2.08	0	0	0	2	3	5	5	4.60
2.09	0	0	0	0	5	5	5	5.00
2.10	0	0	0	2	3	5	5	4.60
2.11	1	1	0	3	0	5	4	3.00
2.12	3	1	1	0	0	5	1	1.60
2.13	0	1	0	2	2	5	5	4.00
2.14	0	1	0	2	2	5	5	4.00
2.15	2	0	0	2	1	5	4	3.00
2.16	1	1	1	2	0	5	4	2.80
2.17	0	0	0	1	4	5	5	4.80
2.18	0	0	0	2	3	5	5	4.60
2.19	0	0	0	1	4	5	5	4.80

Table A3-2

Likert Data Summary - Teacher (continued)

2.20	1	3	1	0	0	5	2	2.00
2.21	0	0	0	2	3	5	5	4.60
2.22	0	0	0	3	2	5	4	4.40
2.23	2	1	1	1	0	5	1	2.20
2.24	0	0	0	1	4	5	5	4.80

Note. For the purpose of calculating the mean and mode the following values were assigned; SD = 1, D = 2, N = 3, A = 4, and SA = 5.

Written Responses

The teachers were asked to respond to eight questions by writing comments on the questionnaire. The questions asked and the responses given are listed below. The responses to the questions are reasonably short and limited in number. As a result it is possible to include all of the responses of all the teachers. The numbers associated with the responses are for reference only and are not associated with any individual teacher.

Question 3: What one feature did you like most about the program?

Responses: 1. The ease of /setting up an exam.

2. The different methods of test questions offered. (True-False, multiple choice, or short answer)
3. The ability to test students on material at a time convenient to all, without having to mark tests or monitor.
4. Test review giving feed back for incorrect response.
5. Students wanted to do well on tests. If they did poorly, they tried very hard to improve their mark.

Question 4: What one feature did you like least about the program?

Responses:
1. A little slow in loading.
2. If you did not pass the test you had to review. No choice.
3. The time required to author a test.
4. Time taken to store questions (speed of program).
5. Creating new units

Question 5: What was the most frequently occurring problem you had in running the program?

Responses:
1. Didn't really have any. Had a problem changing unit menu.
2. No major problems.

3. Forgotten passwords and the time to juggle disks to find password.
4. Saving student marks.

Question 6: What two or three things should be changed to make the program more useful?

Responses:
1. Add a help program option.
2. Grade book to include project marks. Choice of reviewing test or not.
3. Speed.
4. Make the program more user friendly.

Question 7: What testing options beside feedback, mastery level, and access, would you like to see under teacher control?

Responses:
1. The number of test questions each student gets from each topic and for the total test.
2. Questions and what topic areas are in the test.

Question 8: What type of question did you use most?

Responses:
1. Multiple choice
2. Multiple choice and true-false at first then I realized how valuable 3 & 4 opinions (questions per topic?) were when the computer randomly picks questions.
3. Multiple choice and true-false.
4. Multiple choice.
5. Multiple choice.

Question 9: What type of question did you use least?

Responses:
1. Short answer
2. Short answer.
3. Short answer.
4. Short answer.
5. Fill in the blank.

Question 10: Are there any functions that should be added? (If yes please describe)

Responses:
1. Average 3 rather than 2 tests.
2. When generating a written test I would like to be able to change the order that the questions occur on the test as well as previewing the test before printing.
3. Use of IIe "delete key". A printer routine to print a certificate when test completion is successful.

Teacher Interview Responses

The interview questions asked of the teachers and the responses given are listed below. The responses have been edited and those responses dealing with the similar concerns have been combined. Although the responses are numbered, the numbers are for reference only and do not identify any particular individual.

Question 1: What kinds of things gave you problems with the program?

Responses:

1. A number of program errors occurred when the program tried to access the disk drive to retrieve or save information. In some instances data was lost, in others it was necessary to restart the program. The identified problems were applicable to a number of different program functions ranging from reading Unit files to storing student marks.
2. The program's error handling routines did not provide warning messages when the program encountered locked data files or write protected disks. As a result some data was lost.
3. There were some problems in obtaining usable printouts from the test generator. The program would omit parts of some test items.
4. The Restart instructions that appeared, after a student signed off from the program, were not clear.
5. There was some difficulty associated with learning the terminology and learning how to operate the program.
6. Copies of QUIZMASTER were made using the Apple IIc Utilities disk. These copies would not run.

7. Although not strictly related to errors within the program proper, there was some difficulty caused by an incorrectly keyed test item in one of the safety tests provided for use in the field test.

Question 2: Did your students have any problems with the program? If so what problems did they have?

Responses:
1. There were several cases of the program failing to store marks. In some cases the cause for this was unexplained. In other cases the causes were a write protected Student Disk or failure to sign off from the program as directed.

Question 3: What do you think are the greatest weaknesses of the program organization?

Responses:
1. The program is complex enough that a flow chart should be provided showing the paths to the different program functions.
2. The editing features could be more user friendly. Maybe the use of a mouse would help.
3. It is not possible to go directly from the Teacher Disk to the Student Disk or vice versa without restarting the computer using the appropriate disk.

4. The test review is compulsory for all students. Some student find this annoying.
5. The program takes too long to load.

Question 4: What did you like about the program organization?

Responses:
1. The item analysis capability provides feedback on question quality.
2. The ability to generate paper and pencil tests.
3. The program encourages students to do well. They were motivated to try a test a second time to improve their marks.
4. The program is clear and straight forward. It describes what you are doing. It is almost impossible to get lost.
5. QUIZMASTER fits into the Industrial Arts program very well. It is geared for student use.
6. QUIZMASTER makes it easy to make a new test and to edit existing tests.

Question 5: Is the program format (Units, Topics, etc.) effective? If not what would you change?

Responses:
1. It takes some time to become accustomed to the terminology and program organization. After the initial period working with the format is no problem.

2. Yes, but restricted by editing procedures.
3. The program format helps organize thinking.

Question 6: How would you use this program in your classroom?

Responses:
1. The program would be used for general testing in a number of areas. Examples cited are pretesting, safety testing, and testing for theoretical knowledge.
2. The program could be used in conjunction with workbooks to test for knowledge relative to assigned student projects.
3. The tests generated by QUIZMASTER could be used as instructional units.

Question 7: What features of the program are unnecessary?

Responses:
1. No unnecessary features were identified. One teacher, however, did indicate that the item statistics function may not be used.

Question 8: What features would you like to see added?

Responses:
1. The ability to change the order in which the questions are presented.
2. The program could print a certificate or license for a machine.
3. The ability to generate a list of students who have completed a unit.
4. The Apple IIe delete key should be active.

5. The manual could include a brief description of the characteristics of good true-false and multiple choice test items.
6. The ability to store records for up to 30 or 40 students in two separate classes on one Student Disk.
7. Ability to allow two answers to questions.
8. Add the capability of entering project marks to make a complete grade book.
9. Make it possible for the teacher to switch between Student Disk functions and Teacher Disk functions without having to restart the system.
10. Add a new parameter allowing the teacher to set the maximum number of questions per topic.
11. Add a new parameter setting making it possible for the teacher to designate the test review as optional or compulsory.
12. Add a new parameter allowing the teacher to set the maximum number of questions per topic to be selected when assembling a test.

Question 9: Did you find any aspect of the program operation annoying?

Responses:

1. There is a need to know where to go and what to do next. The menu functions were not clear enough for this.
2. Slow operation in loading was annoying.
3. The fact that the program did not always save the student's marks created some problems.
4. It is not possible to renumber or insert topics or to change the topic order.
5. It is not possible to copy questions from one topic to another.
6. When a student fails to sign off properly, the student's marks are accessible to other members of the class.

Question 10: Do you have any comments or suggestions that you would like to make?

Responses:

1. There is a need to give a demonstration to the students. Some students need extra instruction. A flow chart to put on the keyboard would be useful.
2. It would be better to have the program use raw scores rather than percentages.
3. There are some grammatical errors in help screens.
4. The program was easy to handle.

5. There should be provision for a teacher password that will allow the teacher to use the Student Disk.

Student Questionnaire Data Summary

The data obtained from the students participating in the field test is summarized below. For convenience the questionnaire items are reproduced with the data summaries

Likert Scale Data

Students were asked to respond to a number of statements by indicating their relative agreement on a five point rating scale. For ease of reference the statements are reproduced below. The Item numbers associated with the statements did not appear on the questionnaire. They have been added for reference when interpreting the data tables. Table A3-3 summarizes the data collected from the junior high school students, Table A3-4 summarizes the data collected from the senior high school students, and Table A3-5 summarizes the data collected from all students.

(1.01) Once I had the computer disks I could run the program without much help from my teacher.

(1.02) The instructions on the screen were easy to understand.

(1.03) I liked using the computer to take a test.

(1.04) I liked the extra information given during the test review.

(1.05) I prefer writing tests with a paper and pencil.

(1.06) The testing program helped me learn the subject better.

Table A3-3

Likert Scale Data - Junior High Students

Item Summary								
Item #	SD	D	N	A	SA	Item N	Mode	Mean
1.01	12	20	49	47	64	192	5	3.68
1.02	7	14	22	45	108	196	5	4.18
1.03	18	8	14	49	107	196	5	4.11
1.04	36	16	53	39	46	190	3	3.22
1.05	126	26	19	11	14	196	1	1.78
1.06	27	17	68	46	37	195	3	3.25

Note. Number of respondents = 197

For the purpose of calculating the mean and mode the following values were assigned; SD = 1, D = 2, N = 3, A = 4, and SA = 5.

The value of Chi Square was calculated for the response frequencies for each statement on the student questionnaire. The purpose was to determine if the obtained responses were significantly different from a random distribution. With an alpha of 0.01 the critical region for four degrees of freedom starts at a Chi Square value of

13.28. All of the obtained values are above this level. The values range from a low of 19.99 to a high of 267.4

Table A3-4

Likert Scale Data - Senior High Students

Item Summary								
Item #	SD	D	N	A	SA	Item N	Mode	Mean
1.01	2	1	10	15	16	44	5	3.95
1.02	0	0	7	16	24	47	5	4.36
1.03	1	1	3	20	22	47	5	4.29
1.04	1	2	12	12	20	47	5	4.02
1.05	23	14	6	1	3	47	1	1.87
1.06	1	5	13	21	7	47	4	3.59

Note. Number of respondents = 47

For the purpose of calculating the mean and mode the following values were assigned; SD = 1, D = 2, N = 3, A = 4, and SA = 5.

Table A3-5

Likert Scale Data - All Students

Item Summary								
Item #	SD	D	N	A	SA	Item N	Mode	Mean
1.01	14	21	59	62	80	236	5	3.73
1.02	7	14	29	61	132	243	5	4.22
1.03	19	9	17	69	129	243	5	4.15
1.04	37	18	65	51	66	237	5	3.38
1.05	149	40	25	12	17	243	1	1.79
1.06	28	22	81	67	44	242	4	3.31

Note. Number of respondents = 244

For the purpose of calculating the mean and mode the following values were assigned; SD = 1, D = 2, N = 3, A = 4, and SA = 5.

Written Responses

The students responding to the questionnaire were asked to make written comments in response to three questions. A total of 244 students completed the questionnaire. Because of the large number of comments, the responses were categorized by grouping similar statements. The results are listed below.

Question 2: What did you like best about taking a test with a computer?

Interpreted Response	Number of Students
The program was easy to use.	68
No writing involved in completing the test.	63
It was a different, novel or interesting experience.	43
It was faster.	33
Liked the type of questions used.	14
Liked having immediate feedback.	11
Not as much pressure.	10
An opportunity to use a computer.	8
Focused attention on test and/or test items.	5
It was easy to cheat.	5
Helped student to learn the material.	4
An opportunity to review the test.	3
The accuracy of the computer.	1
Individualized the testing process.	1
Good test security. Can't cheat.	1
Total	270

Note: Some students made more than one comment.

Question 3: What did you like least about taking a test with a computer?

Interpreted Response	Number of Students
It took too long for the class to complete the test.	31
Program delay while loading questions.	21
The questions and/or the type of questions.	20
Can't change a response after it is entered.	19
It was a boring experience.	11
Eye strain from green screen or poor quality monitor.	6
Was unfamiliar with computer use.	6
It was difficult to prepare for the test.	4
Resulted in poor retention of material.	3
Student was unable to cheat.	3
No privacy while completing the test.	3
Was forced to complete the test review.	3
There was no test review.	2
The program format.	2
Changing program disks.	2
The student was unable to see the entire test.	2
No print out of test results.	2
The bugs in the program.	2
It randomized the test items.	1
Felt added pressure.	1
Prefer paper and pencil tests.	1

The program gives the student's mark.	1
The experience was too short.	1
The instructions and/or menus were confusing.	1
Dislike using computers.	1
Total	151

Question 4: Please write any comments you may have about using this computer program.

Interpreted Response	Number of Students
The program was a good idea.	36
Liked it. It was interesting.	35
It was easy.	18
It provided an opportunity to use the computer.	4
It was faster.	3
There was no writing involved.	3
Unable to change answer after it was entered.	3
Problems resulting from the use of two disks.	2
The novelty of the experience.	2
The test review was a good experience.	2
Prefer paper and pencil tests.	2
It helps learning.	2
It was too easy.	2
It needs a greater variety of questions.	1
There were annoying bugs in the program.	1
It needs better instructions.	1

Dislike using computers.	1
No privacy during the test.	1
No feedback.	1
Disliked the program.	1
The program was no help at all.	1
It was a boring experience.	1
The review takes too long.	1
Reduced pressure during the test.	1
Computer marks it right away.	1
It needs some graphics.	1
There was a delay while loading questions.	1
Liked the questions.	1
Total	138

Student Interview Response Summary

Sixty four of the students participating in the field test were interviewed by the researcher in an attempt to obtain more complete information. Forty seven of these were junior high school students and seventeen were senior high school students. There was no structured selection process. Rather the students were selected from the classes that were in session at the time of the visit to the school. Students who indicated an unwillingness to be interviewed were not interviewed.

The students were asked to respond to seven different questions. The student's responses were categorized by grouping similar statements. The results are listed below.

Question 1: What kinds of things gave you problems with the program?

Interpreted Response	Number of Students
No problems in operating the program	34
Problems with test content only	12
Unfamiliar with computers	5
Problems with the password	3
Program was slow to retrieve questions	3
Did not like being forced to answer all questions	3
Made typing errors	2
Could not change answer after it is entered	1
Screen bothers eyes	1
Total	64

Question 2: What did you like best about the program?

Interpreted Response	Number of Students
Did not have to write answers	26
Faster or easier to complete test	19
Provides feedback/gives mark right away	9
Novelty or liked to use computer	7
Less pressure when attempting test	7
Not much difference	1
Total	69

Question 3: What did you like least about the program?

Interpreted Response	Number of Students
Nothing	27
Slow to get next question	12
Problems related to test content	10
Cannot go back and change answers	7
Did not provide feedback	2
The program failed to record student marks	2
The test review takes too long	1
Starting up the program	1
Viewing the screen causes eye complaints	1
Easy to lose concentration when using a computer	1
Total	64

Question 4: Did you do the test review?

With the field test version of the program the test review was compulsory unless the program was placed in the pretest mode. Forty one students completed the test review while twenty three students were not given the opportunity.

Question 5: Were the feedback screens and resource lists useful?

Interpreted Response	Number of Students
Program in pretest mode. No feedback screens presented	23
Feedback screens aid the learning process	12
Feedback screens were of some help	10
Did not remember seeing feedback screens	3
Feedback screens were useful (no qualifiers)	6
Feedback screens explained the answers	6
Feedback screens were not useful	1
Feedback screens not useful when question answered correctly	1
Total	62

Question 6: Would you like a last chance to go over the test before it is marked?

Yes	52
No	6
Not sure on no answer	6
Total	64

Question 7: Do you have any comments or suggestions you would like to make?

Interpreted Response	Number of Students
Program easy to use or liked program	13
Print out test items at end of test	2
Need a wider variety of questions to avoid duplication	2
Need more computers	2
Program encourages you to take the test	1
One question at a time is less distracting	1
Use diagrams in question	1
Preferable to written tests	1
Total	23

Appendix 4

QUIZMASTER Program Description

Introduction

This description of QUIZMASTER is intended to provide an overview of the structure and operation of QUIZMASTER as it existed at the time of the field test. Subsequent changes to the program have not changed it's basic organization. The changes that have been made have added depth and versatility to many of the program components or have been changes which simplified the operating procedures.

QUIZMASTER was written to solve some of the testing problems encountered by teachers who, by the nature of their courses, must individualize instruction for their students. A good example of a course requiring individualization is the multiple activity Industrial Arts program. In this program it is normal to find each student responsible for a different learning activity. The use of traditional paper and pencil testing methods often leads to problems in test security and relatively complex test management and record keeping systems. All too frequently these activities occupy valuable time which could be better spent on instruction and laboratory supervision.

QUIZMASTER provides an alternative to paper and pencil testing methods. If the teacher wishes, test administration can be initiated, and carried out, by the student at a time

that the student finds convenient. Each time the program is used a different form of the test is generated.

The QUIZMASTER Unit

With QUIZMASTER, each test is called a Unit. The QUIZMASTER Unit consists of a structured test item pool and a number of special information screens. The way that the program assembles a test and presents it to a student is described in the following paragraphs.

The Test Item Pool

When assembling a test, QUIZMASTER draws its questions from a structured item pool. The structure makes it possible to have, for any given Unit, a number of different tests which will have approximately the same level of difficulty and to cover the same range of learning objectives. To accomplish this, the test item pool is structured on the basis of the Topic. Each Topic represents a different learning objective. A Unit can contain as many as twenty Topics and each Topic can contain four questions. The test item pool can contain four times twenty or eighty questions. The item pool structure for a Unit is illustrated in Figure A4-1.

When QUIZMASTER assembles a test it does so by selecting one question at random from each Topic. A QUIZMASTER Unit can contain a minimum of one Topic and a maximum of twenty Topics. A test assembled by QUIZMASTER,

then, could be as small as one question or be as large as twenty questions.

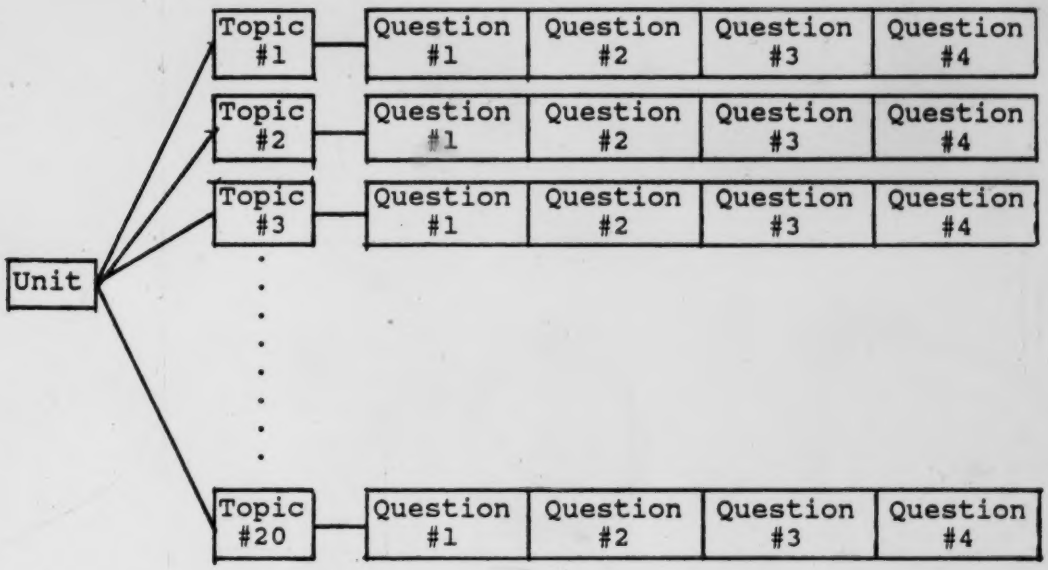

Figure A4-1. Item Pool Structure

Using a structured selection process has the advantage that the teacher knows that every exam presented to the students will cover all of the intended objectives. The random selection process within Topics reduces the likelihood that two students will be given the same test thereby improving the level of test security. The intention is that within any Topic the four questions should all test the same objective and the questions should all have about the same level of difficulty.

Special Information Screens

One of the more important functions of QUIZMASTER is to facilitate the learning process by providing additional information to the student. This information is provided on special screens that may be presented before, during, and after a test. In all cases the use of these screens is optional. If the teacher decides not to use any or all of the special information screens, the student would not be aware of anything missing from the test.

Pretest Information. QUIZMASTER uses two screens to present pretest information to the students: The Unit Objective screen and The Student Introduction screen. The intent of the Unit Objective screen is to make the student aware of the test objectives prior to taking a test. The Student introduction is intended to provide any special instructions for the student. These two screens are available to the student as a help function prior to attempting a Unit. They are presented again when the student attempts a Unit.

Resource Lists. QUIZMASTER has provision for two types of resource lists: The Unit Resource List and the Topic Resource List. These lists provide a means of directing the student to information for further study. The Topic Resource List is presented only when the student answers a question incorrectly. Since a Topic is intended to contain four equivalent questions testing the same

objective there is only provision for one resource list per Topic. The Unit Resource List is displayed after a Unit is completed and only in the event that the student's score on the Unit was less than the preset mastery level.

Feedback. One of the important features of QUIZMASTER is the program's ability to provide the student with a high level of feedback. As part of the entry procedure for every Question the teacher is given the opportunity to enter Mastery Feedback and Non-mastery Feedback that would be unique to that question. Mastery Feedback is presented in the event that a question is answered correctly and Non-mastery Feedback is displayed when the student's response is incorrect. Mastery and Non-mastery Feedback screens are also available at the Unit level. The appropriate Unit feedback screen is presented after a Unit is completed.

Accessing the QUIZMASTER Unit

QUIZMASTER Units are stored on separate data disks. Each data disk will hold two QUIZMASTER Units and the corresponding response frequency data used to perform an item analysis for each Unit. Access to the Unit's contents and response frequency data is controlled by programs which exist on separate Teacher and Student Disks.

Item Analysis Capability

The random selection feature of assembling a test raises questions regarding statistical reliability and test

difficulty. It is desirable that the four questions in a Topic be as closely matched as possible in so far as the difficulty level is concerned. QUIZMASTER facilitates this by accumulating data on each question as it is used. The program will calculate a difficulty index and display response frequencies to allow the teacher to assess the items in any Topic. The teacher may then make changes to the test items as needed.

Student Records

QUIZMASTER's record keeping system allows the teacher to monitor student progress and review the tests as they were written by the students. The program also maintains an ongoing item analysis which permits the teacher to monitor the quality of the test items. QUIZMASTER uses a different Student Disk for each class to be managed. The class list on each student disk will accommodate forty-five students and will maintain marks for as many as thirty different Units.

QUIZMASTER's record keeping system includes a Grade Book and a system for recreating tests. This information is available on the screen or it can be printed on the computer system printer. Printout formats are available that are suitable for teacher records or individual progress reports.

The Student Grade Book

Two marks may be available for each Unit. The first mark is the mark the student earned on the first attempt at the Unit. The second mark is the mark the student earned on the last attempt at the Unit. If a student made only one attempt at a Unit the Grade Book would contain only one mark. If two or more attempts are made the second mark would be the mark earned on the last attempt.

Several report formats are available to the teacher. A class summary is available for each Unit. The Teacher can also request a mark summary for any student. The program is capable of printing out the entire Grade Book or just a final mark summary.

The Grade Book has full editing facilities which allows the teacher to enter or change any student mark.

Re-creating Tests

The random aspect of test assembly makes it impossible to know in advance which questions will be presented to a student. Although this relieves some of the problems of test security it presents other difficulties in diagnosing learning problems. To overcome these difficulties, QUIZMASTER allows the teacher to re-create any test, as it was presented to the student, and to view the student's responses to the objective questions. It is only possible to know whether a student's response was right or wrong for completion type questions.

Controlling the Testing Situation

QUIZMASTER allows the teacher to customize the testing situation in order to suit most conditions. The teacher can control; student access to any Unit, the mastery level for each Unit, the feedback patterns, the resource list display, and the testing mode,.

Unit Access

In order to control class progress the teacher may deny or allow class access to any Unit at any time.

Unit Mastery Level

The teacher may set the mastery level for each Unit. The allowable limits are 1% to 100%.

Feedback Patterns

Changing the feedback mode changes the timing of the Question Mastery and Non-Mastery Feedback. In the immediate mode the appropriate Question feedback statement will be presented immediately after the student enters a response to a question. The deferred mode allows the students to complete the test uninterrupted. In the deferred mode the feedback statements are only presented during the test review.

The test review is offered only when QUIZMASTER is operating in the test mode. Once the test is completed the student is taken through a review. The student will see each question in turn together with the response that was entered. The program will indicate whether the student's

answer was right or wrong. The correct answer is not given. If Mastery or Non-Mastery feedback statements exist then the appropriate feedback statement is displayed. The Topic Resource list is presented only when the student's response to a question is incorrect.

Testing Mode

The teacher may place any Unit into the "test mode" or the "pretest mode". This change alters several factors in the testing situation. In the test mode the student may decline to answer a question and return to it later in the test. In the pretest mode if the student declines to answer a question it is not presented a second time. Regardless of the feedback pattern settings the feedback statements are never displayed in the pretest mode. Similarly the test review is never offered in the pretest mode.

Taking a test

Figure A4-2 illustrates how the various parts fit together when a student elects to take a test. It is important to remember that there are a number of possible variations, depending on the selections made for such things as the test mode, and feedback patterns.

When a student elects to take a test the first screens to be displayed are the Unit Objective and Student Introduction. If these screens were left blank QUIZMASTER will move directly to the first question.

The question administration sequence can have two different patterns depending on the feedback mode selected. These patterns are illustrated in Figure A4-3 and Figure A4-4. If the immediate feedback mode is selected then feedback statements are presented immediately after the student enters a response to each question. In addition the program supplies a message telling the student whether the response was correct or not. The correct answer is not given. If the feedback mode selection is "deferred" the student completes the test uninterrupted.

Upon completion of the test the student's mark is calculated and displayed. The student is then taken through the test review sequence.

Once the test review sequence is completed the appropriate Unit Mastery or Non-Mastery feedback is displayed. The Unit Resource List is only displayed in the event that the student fails to achieve the preset mastery level. In the event that the teacher leaves the Unit Mastery and Non-mastery screens blank QUIZMASTER will supply an appropriate message.

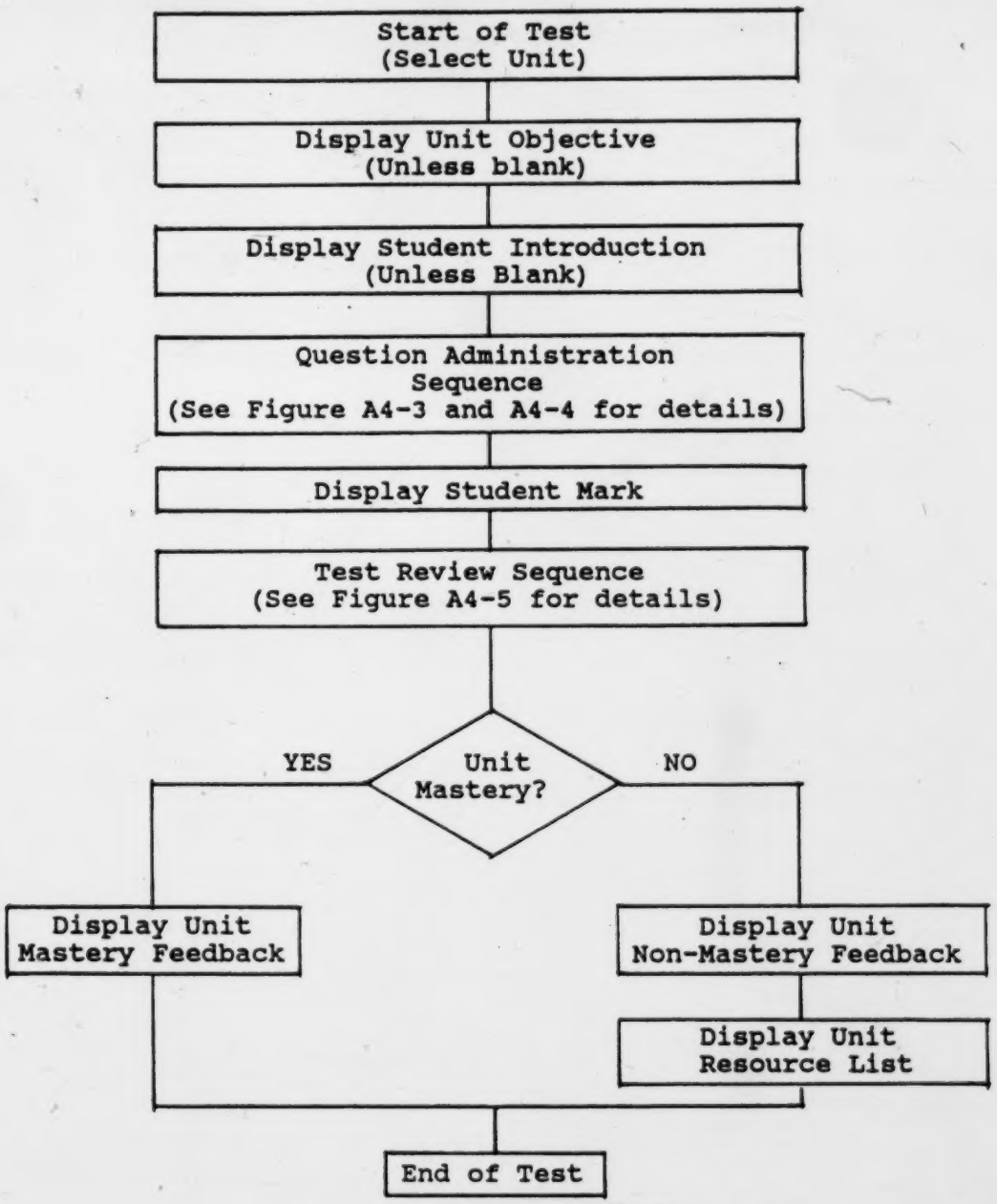

Figure A4-2. Typical test sequence

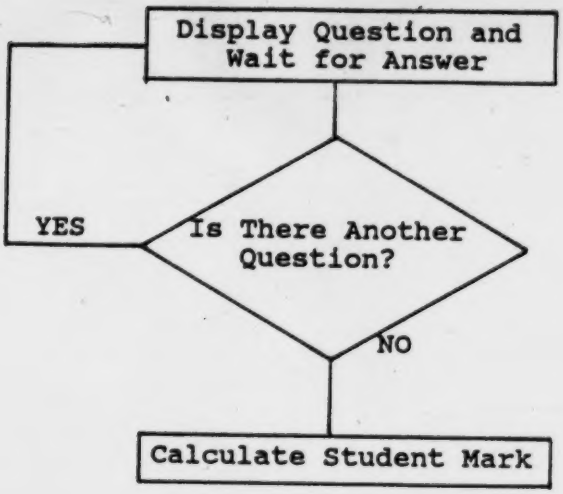

Figure A4-3. Question Administration Sequence - Deferred Feedback Mode

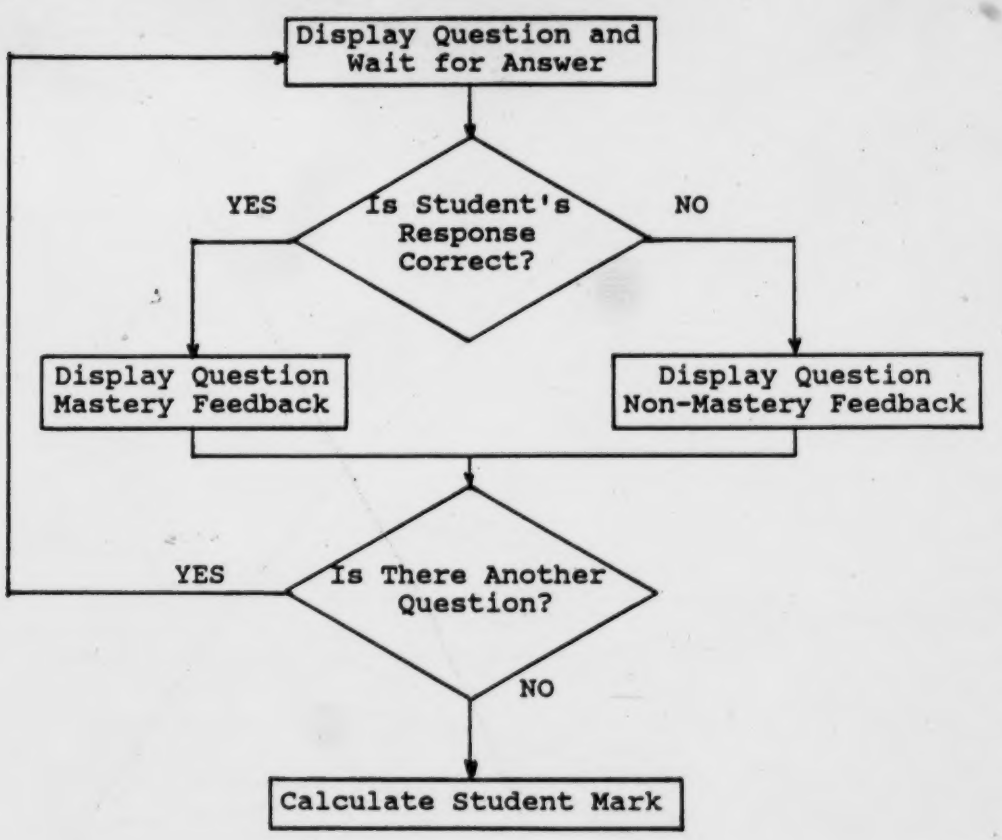

Figure A4-4. Question Administration Sequence - Immediate Feedback Mode

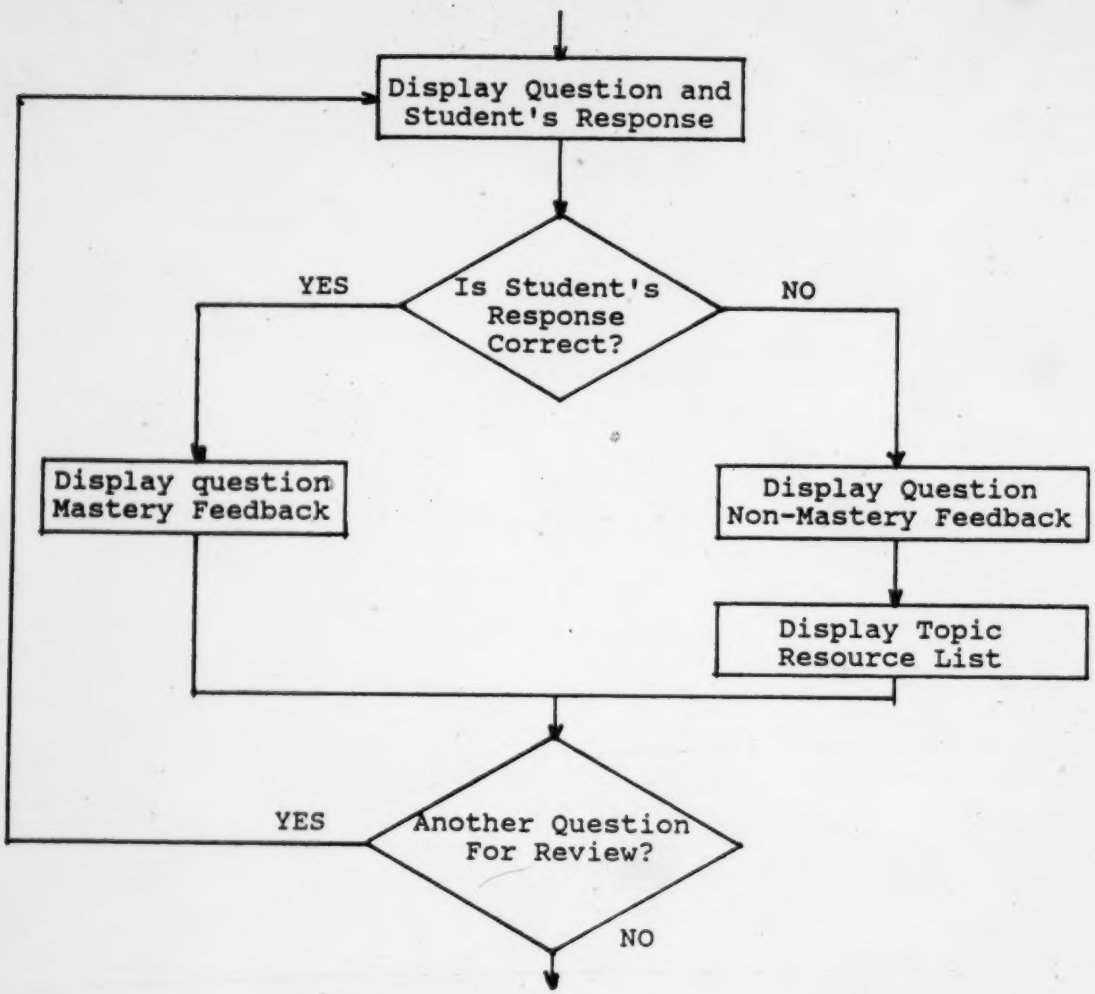

Figure A4-5. Test Review Sequence